Mechanisms behind large-scale inconsistencies between regional and global climate model-based projections over Europe

Ioan Sabin Taranu, Samuel Somot, Antoinette Alias, Julien Boé, Christine Delire

Noname manuscript No.
(will be inserted by the editor)

Mechanisms behind large-scale inconsistencies between regional and global climate model-based projections over Europe

Ioan Sabin Taranu[1] · **Samuel Somot**[2] · **Antoinette Alias**[2] · **Julien Boé**[3] · **Christine Delire**[2]

Received: date / Accepted: date

Abstract Important discrepancies in the large-scale summer climate change projections were recently detected between the global and regional climate models (RCM/GCM) in the EURO-CORDEX ensemble for several variables including surface temperature, total precipitation, and surface solar radiation.

In this study, we use a new experimental framework inspired by the Big-Brother-Little-Brother protocol to explore the mechanisms responsible for generating large-scale discrepancies in future projections between GCM/RCM pairs over Europe in summer. Starting from past and future simulations with a perfect GCM/RCM pair (same resolution, same physics, same forcings), we then disentangle the role of potential sources of GCM/RCM inconsistency by carrying out targeted sensitivity studies.

We show that by following such a perfect approach, it is possible to obtain a GCM/RCM pair without statistically significant inconsistencies in projected climate change. Such discrepancies are mainly generated by differences in aerosols representation and atmospheric physics. The role of plant physiology is limited and unlikely to be the dominant factor in the detected discrepancies. Finally, it is unlikely that the discrepancies in the EURO-CORDEX ensemble projections are a result of the upscaled added value, as we show that the effect of increased resolution is not strong enough and mostly limited to areas with complex topography.

These findings raise important questions about the current practices in regional climate modelling. In the short term, implementing RCM external forcings consistent with the driving GCM can significantly improve the situation at low cost. In the long term, adopting a seamless strategy in developing the GCM/RCM models should be questioned.

Corresponding Author: Ioan Sabin Taranu
E-mail: istaranu2@gmail.com

[1]Vrije Universiteit Brussel, Department of Hydrology and Hydraulic Engineering, Brussels, Belgium
[2]CNRM, Université de Toulouse, Météo-France, CNRS, Toulouse, France
[3]CECI, Université de Toulouse, CERFACS/CNRS, Toulouse, France

Keywords Regional climate modeling · Aerosols · High Resolution · Physics Parameterizations · Added Value · Atmospheric Circulation · Europe · Climate Projection · EURO-CORDEX

1 Introduction

Regional climate models (RCMs) are developed and used to better understand regional climate, climate phenomena and their possible future evolution under climate change. They also serve to meet the needs of the vulnerability, impact, and adaptation community by providing regional climate information to the corresponding climate services.

The methodology involves driving a high-resolution RCM at its lateral boundaries by high-frequency updated 3D atmospheric fields and at its surface boundary by the sea surface temperature and sea ice cover, using information from a lower-resolution global climate model (GCM) [Giorgi and Mearns, 1999, Di Luca et al., 2016]. It is based on the idea that regional-scale climate statistics are conditioned by the interaction between continental-scale atmospheric conditions and regional features such as marginal seas and mountain ranges [von Storch et al., 2000]. Taking this into account, it is commonly assumed that a perfect dynamical downscaling exercise will add consistent high-resolution climate statistics over the region of integration, with no or small impact on the large-scale information of the driving GCM (Tenet 1 and 5) [Laprise et al., 2008].

Today, large ensembles of global and regional climate simulations are available through internationally coordinated projects such as CMIP5 [Taylor et al., 2012], CMIP6 [Eyring et al., 2016], and CORDEX [Giorgi et al., 2009]. Results from CORDEX complement and add value to the CMIP global models, particularly in complex topography zones, coastal areas, small islands, and for extremes - IPCC AR6 Technical Summary [Arias et al., 2021].

RCMs results are not independent of the ones of their driving GCMs, at least for the mid-latitude as, for example, their biases are strongly related to the choice of the driving GCM for most of the climate variables [Vautard et al., 2021]. Therefore, it would be helpful to systematically analyze the RCM data compared to its driving GCM.

However, in practice, the situation is less clear. When climate change signal is analyzed for given GCM/RCM pairs, significant discrepancies are identified even at continental scale. For example, the EURO-CORDEX RCM ensemble projects at the end of the 21st century for the RCP8.5 scenario a weaker summer warming by about 1.5–2 K and a much smaller decrease in precipitation of 5% versus 20% than the respective driving GCM ensemble over continental Europe [Boé et al., 2020]. Having such an inconsistency in the projected climate change between GCM and RCM ensembles is highly confusing for climate data users and can damage the regional and national ability to adapt by reducing the perceived risks. Significant summer large-scale discrepancies in climate change signal over Europe have also been reported with other methodologies and for other variables by various authors [Bartók et al., 2017, Sørland et al., 2018, Schwingshackl et al., 2019, Gutiérrez et al., 2020, Coppola et al., 2021] and assessed in the last IPCC report [Masson-Delmotte et al., 2021]. Similar discrepancies have also been reported for the solar surface radiation for North America [Chen, 2021].

Therefore, it is essential to ask ourselves, are the differences in projected climate change between GCMs/RCMs pairs justified by an *upscaled added-value*, that is to say by a positive large-scale effect of a better representation of the climate fine scales by higher-resolution climate models? The existence of an upscaled added-value would imply that RCM's large-scale projections are better than the GCM's ones. Or are there any flaws in the current practices of dynamical downscaling, which inherently will lead to discrepancies between GCM/RCM projections?

A support of the "added value" explanation is the study of [Sørland et al., 2018] which shows, based on two RCMs that downscaled several GCMs (a total of 14 RCMs simulations), that the RCMs, most of the time, reduce the existing bias between the GCMs and the observations and therefore should be considered as more reliable for future projections. Nevertheless, such an argument should be taken with caution for several reasons, which were also highlighted in the study. Firstly, the improved performance during the observational period will not necessarily translate into a more realistic climate change signal. Secondly, while the GCMs are generally tuned so that the global mean top-of-the-atmosphere energy balance matches observations [Hourdin et al., 2017], RCMs are tuned or calibrated in current-climate simulations driven by reanalysis [Bellprat et al., 2016, Rummukainen, 2016] with the regional distribution of surface temperature and precipitation as the main tuning target. Finally, the mentioned study evaluated the performance of only two RCMs, which may be insufficient to make general conclusion for the entire EURO-CORDEX ensemble.

On the other side, the supposition that the large-scale discrepancy is mainly caused by some flaws in the RCM-based downscaling approach is supported by such studies as [Schwingshackl et al., 2019, Gutiérrez et al., 2020, Boé et al., 2020, Chen, 2021].

RCMs do not generally consider the vegetation response to elevated atmospheric CO2 concentrations, a process that is, however, included in most of the driving GCMs. By activating the so-called stomata effect in the COSMO-CLM regional model, which translates into less plant transpiration at higher CO2 concentrations, it is possible to explain about 67% of the more substantial annual maximum temperature increase in GCMs than RCMs [Schwingshackl et al., 2019].

In [Gutiérrez et al., 2020], the authors have shown that over Europe in the summertime, the difference in projected changes in surface solar radiation between GCM/RCM ensembles is unlikely to be caused by differences in clouds and that the significant factor is the representation of aerosols in the RCM models. It turns out that most of GCMs use time-varying and, often, more complex interactive aerosols schemes, while most RCMs use time-invariant or climatological aerosols.

Finally, in [Boé et al., 2020], the authors noticed significant discrepancies in the summer climate change projections between GCM/RCM ensembles on a continental scale, not only for surface shortwave radiation but also for other variables, such as surface temperature, cloud cover, precipitation, and evapotranspiration. It was suggested that the missing plant CO2 physiological effect, the often-missing evolving aerosol forcings in RCMs, and the lack of ocean-atmosphere coupling in RCMs are some of the potential sources of the observed inconsistencies.

In this context, the goal of our study is to investigate further the role of various mechanisms in the generation of GCM/RCM large-scale discrepancies over Europe in summer. To do so, we do not rely on existing ensembles of opportunity as in previous studies, but we introduce a new methodology based on a perfect-model

approach in the same spirit as the popular Big-Brother/Little-Brother protocol in the RCM literature [Denis et al., 2002b]. To start, we set a GCM/RCM pair sharing the same dynamics, physics, external forcing, and horizontal resolution. This perfect approach will establish a benchmark experiment for the historical and future periods, which will estimate the best achievable coherence with a given pair of GCM and RCM models. Once the capacity to generate a coherent GCM/RCM pair is confirmed, we will deviate from the perfect approach by introducing imperfections such as changing the atmospheric physics, external forcing, or resolution of the regional model. Implementing such an experimental framework is unique and currently possible only for a few groups that follow a seamless development strategy of their GCM/RCM models. This is the case for the CNRM (Centre National de Recherches Météorologiques) laboratory in France, with the ARPEGE/ALADIN models developed seamlessly, meaning that the two models can be run using the same dynamical core, set of external forcing and physical parameterization.

The manuscript is organized as follows: in Sect. 2.1-2.3 we introduce our experimental framework with one benchmark seamless ARPEGE/ALADIN pair, and several sensitivity experiments to isolate the effect of each mechanism on the generation of large-scale discrepancies in projected climate change. Sect. 2.4-2.8 introduces the various methods used to analyze the simulations. Sect. 3 presents the main results, and how they help explain the detected inconsistencies in projected climate change between the GCM/RCM pairs of the EURO-CORDEX ensemble. Finally, in Sect. 4 we provide the conclusions.

2 Data and methods

To explore the mechanisms behind GCM/RCM projections discrepancies, a new experimental framework is proposed by using the advantages of the seamless GCM/RCM pair ARPEGE/ALADIN (see Fig. 1). In the sections below, we explain the advantages of such a framework and how each experiment is configured.

2.1 Reference experiment

We begin by running CNRM-ARPEGE v6.3 (the atmospheric component of CNRM-CM6-1 and CNRM-ESM2-1 in CMIP6) [Roehrig et al., 2020] at a horizontal resolution of roughly 156km, with the SURFEXv8.0 land surface model [Decharme et al., 2019], including the 1D sea-ice module GELATO [Mélia, 2002], and the following external forcings: tropospheric ozone from University of Reading [Li et al., 1995] and 3D time-varying aerosol datasets from LSCE [Szopa et al., 2013]. In addition, the sea surface temperature and sea ice cover is from the historical simulation (r1i1p1 member) carried out with the CMIP5 CNRM-CM5 coupled GCM and the scenario RCP8.5 simulation (r1i1p1 member) for the future period [Voldoire et al., 2013]. The greenhouse gases (GHG) concentrations follow the CMIP5 recommendations for the historical and future periods. In terms of vertical levels, CNRM-ARPEGE v6.3 has 91 levels. A detailed overview of the experimental setup for the reference global and all regional simulations, their configuration, and abbreviations can be seen in Table 1.

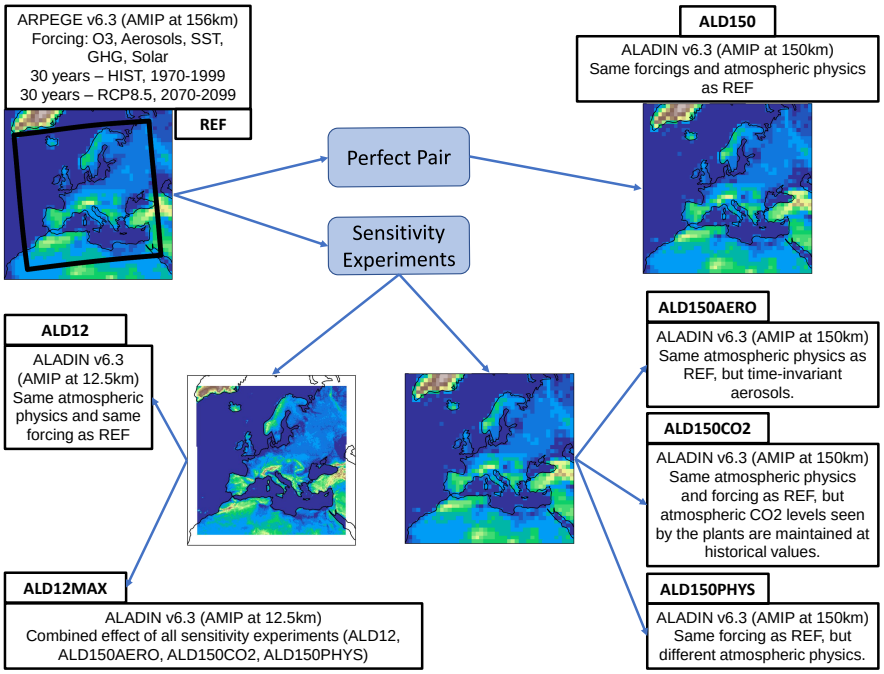

Fig. 1: Schematic representation of the experimental protocol, with a short description of all experiments and their abbreviations. See Table 1 for details on each experiment configuration.

For all the experiments in this study, we ran two 30 year simulations: 1970-1999 for the historical period and 2070-2099 for the RCP8.5 scenario. Two years of spinup were done for each simulation (1968-1969 and 2068-2069), to reach a quasi-equilibrium of the soil moisture.

The atmospheric fields of the reference ARPEGE simulations were then used to generate the boundary conditions necessary to drive the regional model ALADIN every 6 hours. We will refer to this ARPEGE experiment as REF, because it will serve as a reference for the large-scale climate change signal for all regional experiments.

2.2 Seamless regional experiment

In this study, all regional experiments including the seamless experiment are carried out with the CNRM-ALADIN regional climate model version 6.3 [Nabat et al., 2020], driven at the boundaries by the REF atmospheric fields. When developing the ALADIN model, the CNRM follows a seamless development strategy, meaning that identical versions of the ARPEGE/ALADIN models share the same dynamical core, atmospheric physics, and set of forcings (see details in Table 1). Both CNRM-ALADIN and CNRM-ARPEGE models operate with 91 vertical levels, with first level at 6 m above ground and last one at about 82 km [Roehrig et al., 2020].

Simulation	REF	ALD150	ALD12	ALD150AERO	ALD150CO2	ALD150PHYS	ALD12MAX
Model	ARPEGEv6.3	ALADINv6.3	ALADINv6.3	ALADINv6.3	ALADINv6.3	ALADINv6.3	ALADINv6.3
Resolution	156 km	150 km	12.5 km	150 km	150 km	150 km	12.5 km
Aerosols	Szopa	Szopa	Szopa	HIST Szopa for both HIST and RCP8.5 simulations	Szopa	Szopa	HIST Szopa for both HIST and RCP8.5 simulations
Decreased Plant Transpiration at higher CO2 concentrations	Yes	Yes	Yes	Yes	No	Yes	No
Physics	CMIP6	CMIP6	CMIP6	CMIP6	CMIP6	CMIP5	CMIP5
Ozone	Reading	Reading	Reading	Reading	Reading	Reading	Reading
SURFACE	SURFEXv8.0	SURFEXv8.0	SURFEXv8.0	SURFEXv8.0	SURFEXv8.0	SURFEXv8.0	SURFEXv8.0
SST	CMIP5	CMIP5	CMIP5	CMIP5	CMIP5	CMIP5	CMIP5
Ice Module	GELATO	GELATO	GELATO	GELATO	GELATO	GELATO	GELATO
SIC	CMIP5	CMIP5	CMIP5	CMIP5	CMIP5	CMIP5	CMIP5
GHG	CMIP5	CMIP5	CMIP5	CMIP5	CMIP5	CMIP5	CMIP5
Volcano	-	-	-	-	-	-	-

Table 1: Conducted experiments, their abbreviations and configuration description.

In addition to using the same physics and forcings, the seamless CNRM-ALADIN experiment is designed to operate at a horizontal resolution of 150km, almost the same as the driving CNRM-ARPEGE model (156km). By doing so, we created a new variant of the famous Big-Brother/Little-Brother experiment first defined by [Denis et al., 2002b]. Therefore, any deviations from the perfect pair (CNRM-ARPEGE/CNRM-ALADIN) can be classified in the large family of the imperfect Big-Brother/Little-Brother experiments. In the following, we will refer to the seamless regional experiment as ALD150 (see Fig. 1).

It is important to mention, that despite having a very similar resolution and same vertical levels, the position of grid points inside the domain of integration can significantly differ between CNRM-ARPEGE and CNRM-ALADIN models. This is caused by the fact that CNRM-ARPEGE model is running on a reduced Gaussian grid [Roehrig et al., 2020], while CNRM-ALADIN on a Lambert conic conformal projection. This difference in the grid geometries represent a limitation of the proposed experimental framework, and is something to take into account when analyzing the results.

2.3 Sensitivity experiments

Effect of resolution

The ALD12 experiment was designed to quantify the role of upscaled added value on the large-scale discrepancies in projected climate change between regional and the driving global models. ALD12 has identical physics and forcing as the REF and ALD150 experiments (see Table 1), but this time with a resolution of about 12.5 km, the standard EURO-CORDEX resolution. In this context, both ALD12 and ALD150 configurations are eligible for EURO-CORDEX, covering the minimum required domain. In fact, simulation results with the exact same configu-

ration as for ALD12 experiment are accessible in the EURO-CORDEX ensemble [Coppola et al., 2021].

In comparison to the ALD12, the ALD150 experiment domain is larger (see Fig. 1). When designing these experiments, the size of the domain and the number of grid points in the buffer zone were some of the important setups to consider. In the end, because of the coarser resolution, the ALD150 domain was made larger by about 450 km from each side, so in case of regriding, it would be possible to directly compare the results with the higher ALD12 resolution setup. In addition to this, the ALD12 setup have 8 grid points for the buffer zone on each side of the domain, while ALD150 have only 3 grid points. While this setup may not be optimal, and subject to re-evaluation in the future, it gives satisfactory results.

Additionally, it should be mentioned that the physical parameterizations for ALD12, while exactly identical to the ones used in REF, can in itself become a source of discrepancies. This is because the implemented convection scheme in ARPEGE/ALADIN models is resolution-dependent [Guérémy, 2011]. To compensate for the resolution-dependence, tuning may be used when the spatial resolution is modified.

In this study, no re-tuning was done for the ALD12 setup. Therefore, the resolution-dependence of the convection scheme may represent a potential limitation in the capacity to fairly compare the ALD150 and ALD12 simulations in order to discriminate the role of upscaled added value. It is worth noting that such challenges always arise, in any resolution added-value studies, when comparing low and high-resolution climate models (GCM or RCM).

Effect of time-invariant aerosols

As previously mentioned, both REF and ALD150 use time-varying aerosols following the LSCE dataset [Szopa et al., 2013]. The five aerosols represented are dust, sea salt, sulfate, black carbon, and organic matter. The ALD150AERO experiment was performed to quantify the effect of time-invariant aerosols on the GCM/RCM inconsistency. The only difference with the ALD150 experiment is that for the RCP8.5 period of 2068-2099, the same aerosol concentrations were used as in the historical period 1968-1999 (see Table 1). The correspondence is done month by month, such that for example the aerosols' concentration in March 2075 will be the same as in March 1975. This will assure an identical representation of aerosols between the two periods, in which case the difference in the climate change signal between ALD150 and ALD150AERO can be connected directly to the role of time-invariant aerosols.

Plant physiological CO2 effect

In their study, Schwingshackl et al. considered the effect of closing stomata at higher CO2 concentrations, which causes an increase in surface temperature by reducing evapotranspiration. In the case of the ALADIN model, it is possible to reproduce this experiment, as ALADIN is one of the few RCMs with implemented plant physiological response to CO2 variations.

In the REF experiment, the CO2 "antitranspiration" effect [Séférian et al., 2019] is already active. Therefore, to test the effect of absent CO2 related increase in plants water use efficiency, we designed the ALD150CO2 experiment. In the

ALD150CO2 setup, the CO2 concentration seen by the land-surface model is kept constant at the 1984 historical level (about 345 ppm) for both historical and RCP8.5 simulations. At the same time, the atmospheric CO2 concentrations continues to rise following the RCP8.5 scenario. Consequently, through its control of leaf transpiration, the higher future stomatal conductance in ALD150CO2 is expected to cause higher latent heat flux and affect the surface energy balance.

Effect of GCM/RCM inconsistencies in atmospheric physics

In the first generation of EURO-CORDEX, the selection criteria for the GCM/RCM pairs were not explicitly defined. The end result is a large ensemble of opportunity with the GCMs and RCMs models available at the time, trying to maximize the number of GCM/RCM combinations. Consequently, the GCMs and the driven regional models often have significant differences in modeling atmospheric and land-surface processes.

To the best of our knowledge, little is yet known about how the large-scale discrepancies between the GCM/RCM ensembles are connected to differences in the physical package between GCM/RCM pairs. The ALD150PHYS experiment was done to quantify the potential role of such structural differences. While REF and ALD150 simulations use the same atmospheric physics (called CMIP6 in Table 1), the ALD150PHYS uses the atmospheric physics as it was during the CMIP5 project (similar to the one used in [Colin et al., 2010, Bador et al., 2017]). The main differences between the two model versions are in their cloud, deep convection, and turbulence schemes [Daniel et al., 2019]. At the same time, they still share the same radiative transfer model for both short and long waves and the same surface scheme.

It is worth mentioning that we are proposing here only one among many plausible tests for checking the effects of the GCM/RCM inconsistencies in physical parameterization.

Combined effect

We will see throughout the other experiments how differences in resolution, aerosol forcing, plant physiology, and physics cause discrepancies between the GCM and driven RCM. Furthermore, we would like to know if the contributions of these effects are additive and what is the maximum achievable discrepancy that we can obtain in our experimental framework. To do so, we did the ALD12MAX experiment, where we combined the effect of the resolution, aerosols, plants physiology, and physics. It should be mentioned that this type of simulation is relatively close to what is happening in a typical CORDEX experiment and have the potential to help explain the current discrepancies between GCMs/RCMs ensembles.

2.4 Methods to quantify the GCM/RCM large-scale discrepancies

The REF and all ALD12 experiments results are first interpolated on the native grid of ALD150 using second-order conservative remapping [Schulzweida, 2021]. The resulting files are in the Lambert Conformal projection, with a resolution of 150 km. It is essential to mention that all comparisons and analyses in this

study are made at this low resolution of 150km, as we are interested only in the large-scale discrepancies.

Spatial averages over continental Europe are computed for the land points between 42° N, 52° N, -5° E, and 30° E (see red box in Fig. 2) to quantify the mean large-scale discrepancies. The land points have a fraction of land greater than 0.75 after interpolating the REF land mask on the ALD150 grid. The same definitions of land points are used for the RCM experiments. The same method was used in [Boé et al., 2020] and was chosen to compare our results with this previous study.

For the maps, the discrepancy is computed as the difference between the climate change signal of a particular ALADIN simulation and REF ($\Delta\Delta = (ALD_{RCP85}$-$ALD_{HIST})$ - $(REF_{RCP85}$-$REF_{HIST})$). For the historical period (only for Sect. 3.2), the discrepancy is computed between the historical climatologies of ALD150 and REF ($\Delta = ALD_{HIST}$ - REF_{HIST}). This operation order, ALD-REF, is followed for all discrepancies maps in this study.

For the historical period, a paired two-sided Student's t-test was used at each grid point to assess the statistical significance of the differences between REF/ALADIN ($p < 0.05$), where each pair is the difference of the summer mean for the given year.

For the differences in projected climate change, we first subtract the mean summer climatology for the entire period (1970-1999) from each year's summer mean for the RCP8.5 simulation (2070-2099) for both REF and ALD experiments individually. This procedure will provide for each experiment a series of 30 Δ values, corresponding to the differences in mean summer conditions for a given year in RCP8.5 simulation, in comparison to mean summer conditions for the full historical period. The paired two-sided Student's t-test is then applied to the differences of these series between the ALD and REF experiments.

Conducting significance tests separately at each grid point can increase the probability of false rejection of the null hypothesis [Wilks, 2016]. We, therefore, control the false discovery rate with the approach described in [Wilks, 2006], using an α_{FDR} of 0.05, assuring a more robust statistical test. Grid meshes with differences considered significant are represented with hatching on the maps.

2.5 Spatial correlation analysis

The spatial structure of detected discrepancies is often complex, and it can be challenging to track down how each variable connects to others. As a tool for tackling this issue, a Pearson-based pairwise spatial correlation analysis was implemented between prognostic variables of interest (e.g., surface temperature) and potential sources of these discrepancies (e.g., differences in topography map between the GCM and the RCM). In Sect. 3.2, the spatial correlation analysis is used to evaluate the potential role of differences in topography and coastlines definition between the ALD150 and driving REF in generating large-scale discrepancies.

Even at similar horizontal resolutions (156 vs. 150 km), significant differences in the topography and coastlines definition result from different native grid geometries of ARPEGE and ALADIN. At the same time, the effect of such differences on surface representation is expected to be stronger in places with high topography and coastlines, where a slight displacement in the gridcell position can lead to

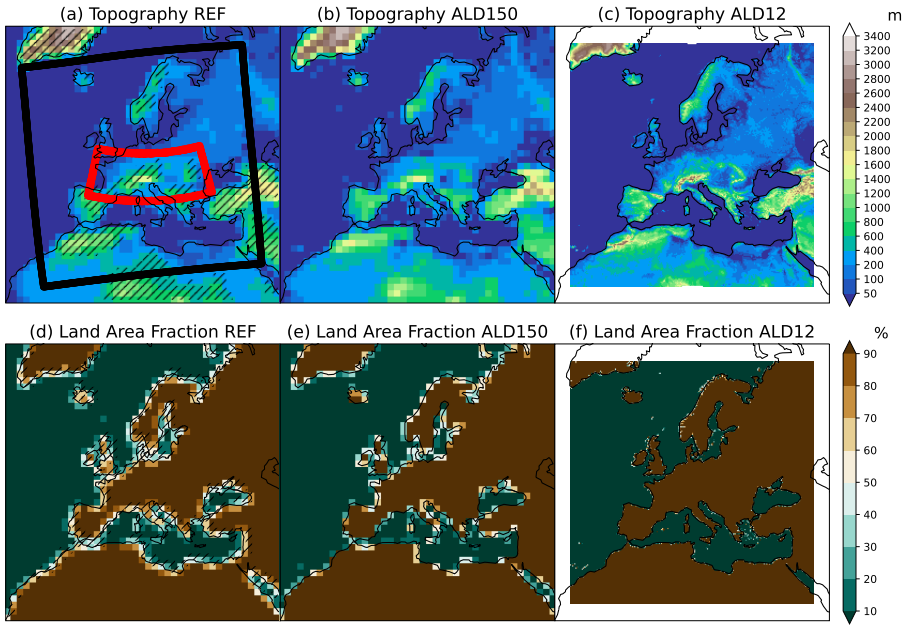

Fig. 2: Topography and land area fraction maps for REF, ALD150, and ALD12. The REF is presented after conservative remapping on the ALADIN Lambert Conformal native grid at a horizontal resolution of 150 km, while ALD12 on its original grid. The dashed lines in (a) show regions where topography equals 500 m or higher. The dashed lines in (d) show grid points with a fraction of land between 20% and 75%. The red box in (a) represents the area over continental Europe over which the mean GCM/RCM projection discrepancy is computed (see Sect 2.6). The black box in (a) represents the official EURO-CORDEX minimum domain over which the fraction of incoherent grid points is computed (see Sect. 2.7).

important changes in altitude and ocean to land fraction. Therefore, to correctly estimate the magnitude of the correlation, it is essential to consider only the relevant grid points. Consequently, the spatial correlation was computed over two masks, both based on the interpolated REF land mask. For the role of topography differences, grid points with an altitude of 500 m and higher were used (hatched areas in Fig. 2.a). For the role of coastline definition, grid points with a fraction of land between 20% and 75% were selected (hatched areas in Fig. 2.d).

2.6 Box plots for mean discrepancies

To quantify how much of the mean discrepancies in the GCM/RCM pairs of the 12-km RCP8.5 EURO-CORDEX ensemble are reproduced by each experiment in the study, we use box plots featuring the mean, median, interquartile range, the whiskers, and the outliers.

Firstly, the spatial average of discrepancies is computed for each GCM/RCM pair available in the EURO-CORDEX ensemble over the red box in Fig. 2 for

the six variables, evapotranspiration, precipitation, 2-m surface temperature, surface downwelling shortwave radiation, pressure at sea level and total cloud cover (EVSPSBL, PR, TAS, RSDS, PSL and CLT) using the approach described in Sect. 2.4. An ensemble of 45 GCM/RCM pairs is obtained in consequence, for which the data is available for all required variables. If we consider the GCM/RCM pairs that have at least one of the required variables, we get 48 pairs (see all ensemble members in Table SM. 1). For the generation of the plots, all the data is used, including the 3 GCM/RCM pairs with incomplete set of variables.

Similarly, the spatial average of discrepancies is computed for each of our experiments. This allows us to assess the magnitude and sign of mean discrepancies for each experiment across all variables of interest, and also place these results in the context of the EURO-CORDEX ensemble.

2.7 Histogram plots with fraction of incoherent grid points

While the box plots with mean discrepancies are excellent to summarize the mean magnitude of the GCM/RCM projection discrepancies over a large area of interest, as well as to put our results in the context of previous studies [Boé et al., 2020], it is easy to underestimate the degree of inconsistency, because of canceling positive and negative contributions.

This is why an additional metric is proposed based on the fraction of grid points of the EURO-CORDEX minimum domain (black box, Fig. 2) where a statistically significant difference is observed (see Sect. 2.4 for details on how statistically significant points are identified). Since the EURO-CORDEX focus is mainly applications on land, only land points are considered, which means grid points with a fraction of land higher than 75%. While such an index does not provide information on the magnitude of the discrepancy, it has the advantage to be easily generalizable to any new simulation or other CORDEX domains and will allow comparison with future studies. Together, the box plots and the fractions of incoherent grid points allows obtaining a more complete picture of the GCM/RCM discrepancy.

Such fractions of incoherent grid points were computed for each REF/ALD experiment and for each GCM/RCM pair in the EURO-CORDEX ensemble, which allowed to compare the level of consistencies between the two datasets.

2.8 Impact of each mechanism on the GCM/RCM large scale dynamics

It can be expected that a GCM/RCM pair sharing the same dynamical core, resolution, physical package, and forcings will be very coherent in terms of large-scale dynamics. We can also assume that the RCM large-scale flow inside the domain may deviate from the GCM large-scale flow if inconsistencies are introduced in the RCM, as in our sensitivity experiments. In order to verify the first hypothesis and to measure the effects of GCM/RCM inconsistencies on large-scale dynamics, a metric based on the work of [von Storch et al., 2000] was implemented.

Firstly, the 6-hourly instantaneous wind data at the 500 hPa atmospheric level of each experiment was interpolated on the grid of ALD150. Since the ALD12 domain is slightly smaller than the ALD150 domain (see Fig. 2.b-c), after the interpolation, the ALD150 and REF wind data outside the ALD12 domain is

ignored. This procedure is essential because assuring the same number of used grid points in each dataset impose an identical set of atmospheric modes and, theoretically, the same amount of energy contained inside the domain. It should be mentioned that the same number of grid points was imposed only to analyze the dynamics. For all other maps and analyses, all information was preserved.

Then the interpolated data was used to compute the kinetic energy power spectrum (KE) for each 6-hour interval for both Historical and RCP8.5 periods (example of mean spectrum in Fig. 3 left). The problem of the aperiodic structure of atmospheric fields in the limited area models is overcome by using the 2D-DCT (Discrete Cosine Transform) technique [Denis et al., 2002a].

Finally, the similarity between GCM and RCM dynamics is computed for each time interval using:

$$P(t) = 1 - \frac{\langle [\Psi_{ALD}(t) - \Psi_{REF}(t)]^2 \rangle}{\langle \Psi_{ALD}(t)^2 \rangle} \tag{1}$$

Where $\Psi_{ALD}(t)$ and $\Psi_{REF}(t)$ are the kinetic energy spatial variance at time t for the analyzed ALD and REF experiment, respectively, for a given wavelength λ. The $\langle . \rangle$ is the average of the spectrum over the large-scale modes. In our case, the large atmospheric modes were defined as the ones having their wavelength $\lambda \in [1000, 6000]$ km (blue box in Fig. 3 left). This choice is explained by the fact that the largest observed atmospheric mode is of the size of the domain, which is about 6000 km. Meanwhile, the shortest wavelength associated with large scales, 1000 km, was chosen as a good estimate for the Rossby radius of deformation at mid-latitudes. When $P(t)$ is close to 1 (see Fig. 3 right), it implies a high similarity between the REF and ALD large-scale dynamics. More $P(t)$ deviates from 1, and less similar are the large scale dynamics between the GCM and driven RCM for a given time t. $P(t)$ will allow measuring the impact of the various sources of GCM/RCM structural inconsistencies on the dynamical inconsistencies. It should be mentioned that Fig. 3 is given here as support to understand the methodology, but the full analysis is presented only in Sect. 3.10.

3 Results

3.1 Projected climate change

Figure 4 shows the summer mean climate change across multiple variables between end of century, 2070-2099, and historical period, 1970-1999, in RCP8.5 scenario as projected by the REF experiment.

The observed change patterns share many similarities with the results for the CMIP5 and CMIP6 GCMs ensembles [Coppola et al., 2021, Gutiérrez et al., 2021]. In Europe, the projected warming shows a North-South gradient (Fig. 4.a), with hot spots over the Iberian Peninsula, Mediterranean region, Balkans and Caucasus. The weakest warming is observed over the Northeast Atlantic Ocean, while the strongest in the Arctic Ocean, likely due to the recession of summer sea ice.

In line with the results from CMIP5 GCMs [Bartók et al., 2017, Boé et al., 2020], solar brightening is observed over most of the European domain (Fig. 4.e). The highest increase in downwelling shortwave surface radiation (RSDS) is observed

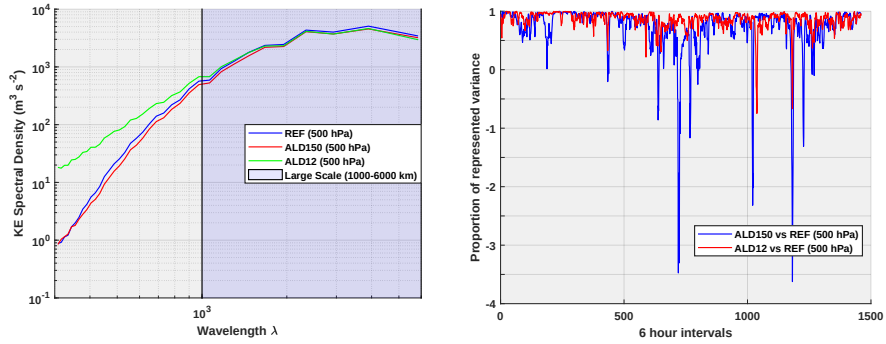

Fig. 3: (left) mean kinetic energy spectra for instantaneous 6-hourly horizontal wind for the year 1970 computed at 500 hPa for the three main configurations used in this study, REF, ALD150 and ALD12. The blue region encapsulates the atmospheric modes corresponding to large-scale dynamics (more details in Sect. 2.8). (right) similarity index for large scale dynamics, $P(t)$, computed every 6-hour for the year 1970, for ALD150 vs REF and ALD12 vs REF (more details in Sect. 2.8).

over Central Europe, which can exceed 30 W/m2 over some grid points over Germany. The increase in the solar energy available at the surface is partially conditioned by the change in cloud cover, but mostly due to projected changes in aerosols concentration, dominated by the decrease of sulfates AOD (aerosol optical depth) [Szopa et al., 2013, Gutiérrez et al., 2020].

Projections show an important decrease in precipitation over most of the central and southern Europe, with a maximum in the Mediterranean region, Spain and France (Fig. 4.c). At the same time, a small increase in precipitation is observed over Scandinavia. These results are in line with EURO-CORDEX, CMIP5 and CMIP6 ensemble projections [Coppola et al., 2021]. The drier conditions in southern Europe cause a reduction of evapotranspiration (EVSPSBL) due to soil water limitation, while an increase is observed over the northern Europe, fuelled by available soil moisture and increased temperatures at the surface.

3.2 Benchmark experiment

Discrepancies during the historical period

Even if ALD150 and REF share the same dynamical core, atmospheric physics, set of forcings and almost the same horizontal resolution, their summer 30-year climatologies over the historical period show significant differences over large parts of the domain (Fig. 5). There are several reasons why this is happening, which are discussed in the following.

The inconsistency pattern that is observed at the far North and East outflow of the domain (see, for example, East boundary in Fig. 5.a) is likely inherent to the one way coupling approach between GCM and RCM. According to our current understanding, the small GCM-RCM discrepancy in the synoptic circulation propagate throughout the domain and eventually grow to divert the RCM large scale

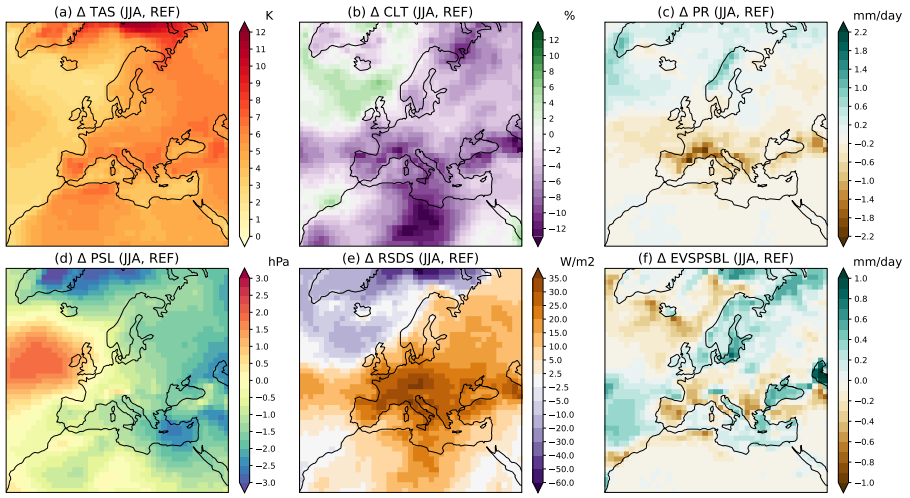

Fig. 4: Mean changes in summer surface temperature (TAS, K), total cloud cover (CLT, %), total precipitation (PR, mm/day), pressure at sea level (PSL, hPa), surface downwelling shortwave radiation (RSDS, W/m2) and evapotranspiration (EVSPSBL, mm/day) as projected by the REF simulation between the RCP8.5 end of century (2070-2099) and historical period (1970-1999).

circulation from the driving model. This creates physical incompatibilities between the RCM and GCM fields at the outflow boundaries, where boundary conditions are overspecified [Staniforth, 1997]. The Davies boundary conditions damp relatively small-scale disturbances, smoothing the fields near the lateral boundaries. However, they cannot handle long waves that reflect from the sponge layer along the boundaries. These reflecting waves interfere with and distort the synoptic circulation across the grid, overwhelming the supply of correct information entering through the inflow boundaries [Miguez-Macho et al., 2004].

For surface temperature, the most important discrepancies are observed in areas with complex topography (see Fig. 5.a in the Alps, Pyrenees, Carpathian, Caucasian regions) and near the coastlines. We assume that these inconsistencies occur because of differences in the surface representation between the ARPEGE and ALADIN models. Despite a similar horizontal resolution (156km for REF and 150km for ALD150), the two models operate on different native grids (Gaussian for REF and Lambert Conformal for ALD150). Even in the scenario of completely identical horizontal resolution, the position of each gridcell would differ between ARPEGE and ALADIN. This difference in geometry translates into differences in physiography between the two models. This effect is most significant in areas with high relief and near the coastlines (Fig.6).

To quantify the effect of physiography differences on the REF/ALD150 mean historical discrepancies, a spatial correlation analysis was done (see methods in Sect. 2.5). The most substantial source of inconsistency for the historical experiment is the difference in topography between REF and ALD150, which is responsible for a significant part of the discrepancies, explaining about 70% of the differences in surface temperature and 25% in cloud cover over the regions with

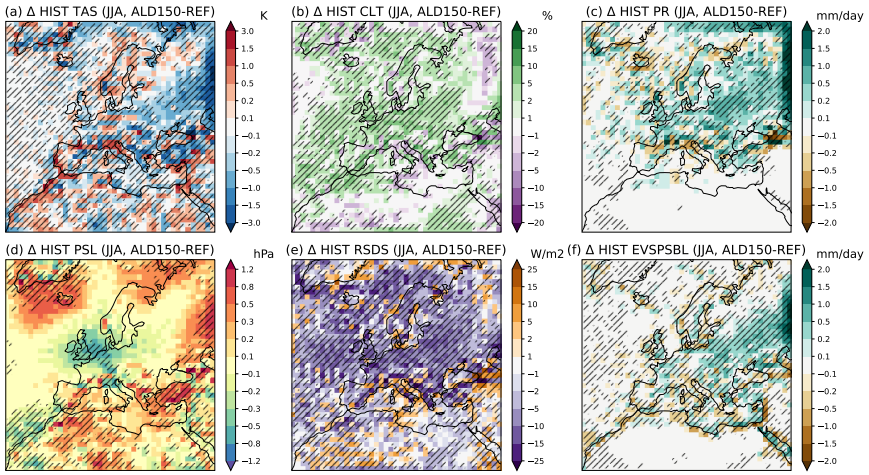

Fig. 5: Mean summer discrepancies between ALD150 and REF for the period 1970-1999 are computed as the difference between the seasonal means of ALD150-REF. Dashed areas are statistically significant (see details in Sect. 2.4). The six variables for which the differences are plotted are surface temperature (TAS), surface downwelling shortwave radiation (RSDS), total cloud cover (CLT), precipitation (PR), evapotranspiration (EVSPSBL), and pressure at sea level (PSL).

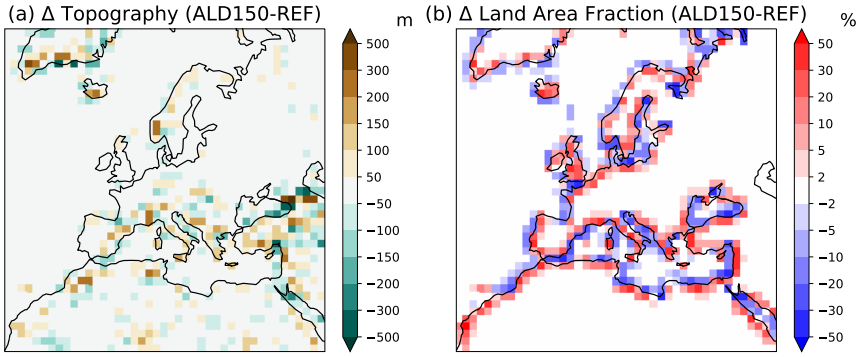

Fig. 6: Differences in topography and land area fraction between ALD150 and REF. The differences are represented on the ALD150 native grid, Lambert Conformal, at a resolution of 150 km.

high topography (see Fig. 7.a, the explained variance is calculated using R^2). As can be expected, a positive difference in relief in ALADIN vs. REF over the same grid point (see Fig. 6.a) translates into a relative cooling of surface temperature due to the higher altitude and a potential increase in total cloud fraction due to the increased saturation chance of moist air subject to orographic lifts. At the same time, the local cooling at the surface may reduce the convection intensity and have an opposite effect, which may explain why the correlation for cloud cover with differences in topography is not as significant as for surface temperature. In-

directly, due to changes in CLT, the differences in relief are also responsible for discrepancies in surface downwelling shortwave radiation (RSDS). It is essential to notice the role of topography differences in generating discrepancies in pressure at sea level (PSL). These regions may then contribute to deviations of the synoptic scales inside the domain when compared to the driving GCM (not shown) and therefore drive other surface variable differences.

The role of coastlines is more modest than that of topography (see Fig. 7.b), explaining about 20% of discrepancies in the total cloud cover and surface solar radiation and 10% in evaporation over the grid points qualified as a coastline. The primary mechanism through which coastlines difference generate inconsistencies is through changes in the properties of the surface. A positive difference in the land area fraction near the coastlines means that a specific grid point in ALD150 has a lower ocean fraction than the respective grid point in REF, resulting in lower evaporation.

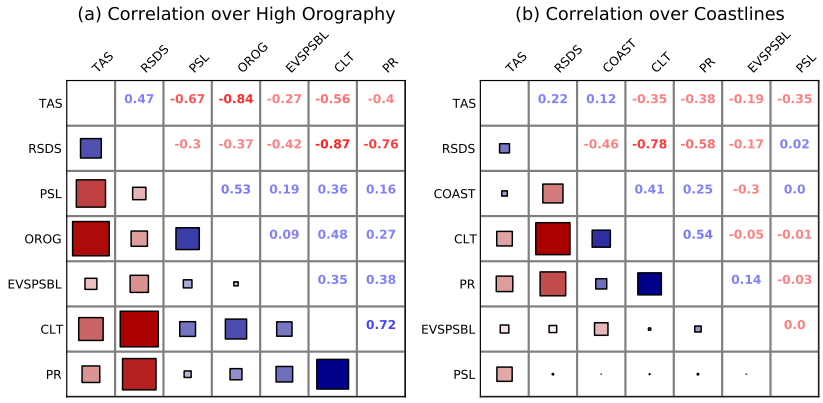

Fig. 7: Pairwise Pearson spatial correlation for historical discrepancies between ALD150 and REF (see Fig.5). To isolate the potential role of differences in topography (OROG) and coastlines (COAST), the spatial correlation was computed over two distinct regions (see details in Sect.2.5). Negative correlations are given in red, and positive correlations in blue. The numbers indicate the correlation coefficients, R, between each pair. The size of the squares indicates the relative magnitude of the correlations and helps navigate the chart faster. The six variables for which pairwise correlation is computed are surface temperature, surface downwelling shortwave radiation, total cloud cover, pressure at sea level, precipitation, and evapotranspiration (TAS, RSDS, CLT, PSL, PR, and EVSPSBL).

The significant discrepancies in total cloud cover over a large part of the domain, including the ocean (see Fig. 5.c), is not entirely understood.

The rest of the discrepancies are probably caused by a combination of factors: a slight difference in resolution between REF and ALD150 (156 vs. 150 km); the difference of gridcells positions affecting not only the surface but the entire column position, and as a consequence, potentially the circulation inside the RCM domain;

small differences in forcing fields (aerosols and ozone) and sea-ice which will be affected by the difference in grid geometries similarly to the relief.

Finally, while the inconsistencies for the historical period seem very strong, they are much lower for REF/ALD150 pair than for most EURO-CORDEX GCM/RCM pairs (see Fig. SM. 4-5).

Discrepancies in projected climate change

While important discrepancies are observed for the historical period, the situation is better when we look at differences in projected climate change. A high degree of coherence in projected climate changes is achieved with the REF/ALD150 pair (Fig. 8). This coherence extends to all the variables.

It can be surprising to obtain such coherence in projected climate changes, while substantial inconsistencies exists in the historical period. However, it is likely because the differences in geometry remain the same throughout the HIST and RCP8.5 simulations. As a result, their effect is systematic and cancels out when the climate change signal is computed.

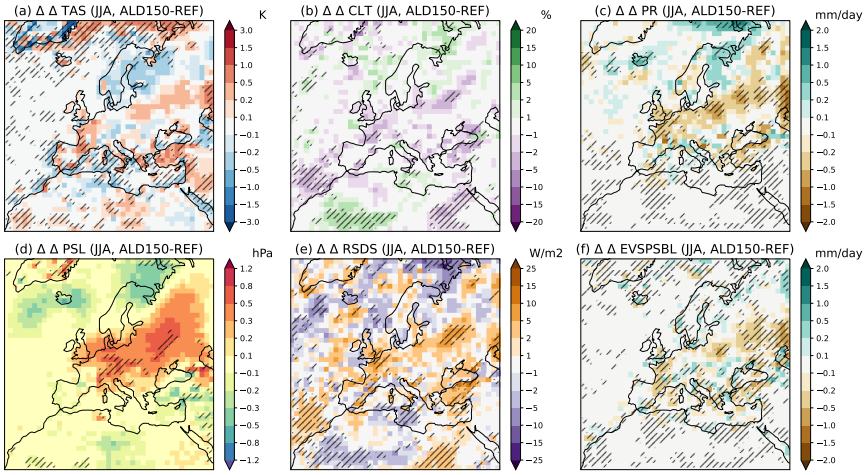

Fig. 8: Mean summer discrepancies in projected climate change signal between ALD150 and REF. The $\Delta\Delta$ means that we first compute the mean projected change for each model between the RCP8.5 (2070-2099) and HIST (1970-1999) simulations, and then we make the difference of resulting Δ's as ΔALD150- ΔREF. Dashed areas are statistically significant (see details in Sect. 2.4).

There are a few points that remain significant, and they are often located near the coastlines (see, for example, Iberian Peninsula coastlines in Fig. 8.a) and in the regions with sea-ice (see, for example, the Greenland Sea in Fig. 8.a). In the previous section, we discussed the role of coastline and sea-ice definition in generating inconsistencies in the historical period. However, these discrepancies are not canceling out when we compare the projected changes ($\Delta\Delta$), likely because the sea surface temperature and sea-ice cover are data driven (SST and SIC obtained

from a previous coupled CNRM-CM5 simulation, see Sect. 2.1). Because of this, when there are differences in ocean to land fraction or sea ice cover between the REF and ALD, the experienced climate change signal by the concerned grids will be perturbed more or less in a non-systematic way by the signal of SST and SIC variables from a different simulation (CNRM-CM5, see details in Sec. 2.1).

3.3 Role of resolution

In terms of the projected change in surface temperature, almost no statistically significant differences are observed over Europe for the REF/ALD12 experiment (see Fig. 9). However, more precipitation and higher total cloud cover are projected by ALD12 over the Alps and Norway. As a response, there is a reduction in shortwave energy available at the surface and an increase in evapotranspiration. These increased horizontal resolution effects agree with the expected added value from improved physiography, even after interpolation to a coarser grid [Giorgi et al., 2016].

A substantial reduction in the total cloud cover and precipitation changes is observed over the Mediterranean Sea for the ALD12 compared to REF (see Fig. 9). This can be perhaps connected to the discrepancy in the projected change of sea-level pressure in the Mediterranean. However, an alternative explanation is that the changes in wind spatial and temporal patterns at higher resolution will influence the Mediterranean air-sea exchanges.

The generality of our findings for the REF/ALD12 pair is difficult to predict. This is partly because the convection scheme implemented in ARPEGE and ALADIN 6.3 depends on the horizontal resolution [Guérémy, 2011]. For this study, the strategy was to maintain the original physical parameterization for both REF and ALD12. Therefore, no tuning was applied to adapt for the increased resolution. Nevertheless, since the discrepancies between REF/ALD12 are surprisingly substantial over the Mediterranean Sea, this problem will be investigated in the future.

3.4 Role of time-invariant aerosols

In the ALD150AERO (see Sect. 2.3), the concentration of aerosols in the RCP8.5 experiment is kept the same as for the historical experiment. This allows isolating the effect of time-invariant aerosols on generating discrepancies between GCM/RCM pairs, as it is done in most of CORDEX simulations so far [Gutiérrez et al., 2020]. As a result, an important difference in the projected change of surface solar radiation (RSDS) is observed (see Fig. 10.e). This difference cannot be explained by changes in total cloud cover, which are similar to the benchmark experiment.

In a time-varying aerosols experiment (i.e., REF), the concentration of sulfates decreases over time due to their short lifetime and expected decrease in emissions due to stricter pollution control measures [Van Vuuren et al., 2011]. Consequently, in REF and other GCMs, a future solar brightening is expected [Gutiérrez et al., 2020, Boé et al., 2020, Chen, 2021]. In the case of REF, the summer solar brightening is about 8.3 W/m^2 over the entire ALD150 domain, with its maximum over Europe, which can achieve up to 40 W/m^2 (see Fig. 4). However, this is not the case for the ALD150AERO, with no summer solar brightening

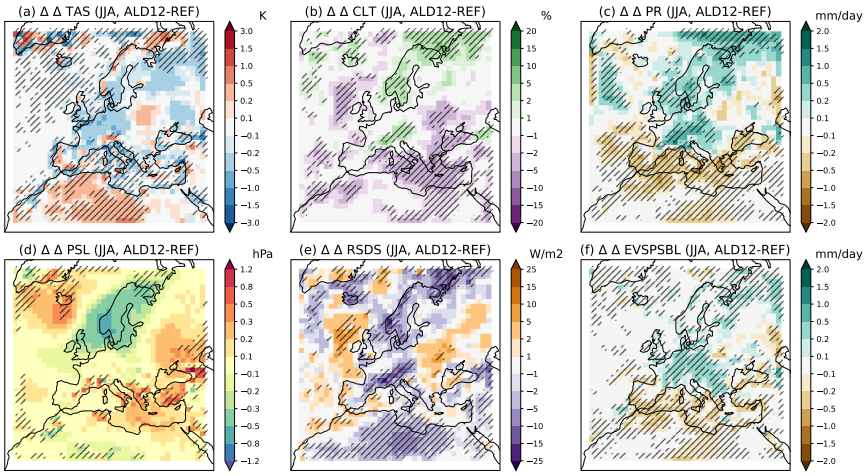

Fig. 9: Mean summer discrepancies in projected climate change signal between ALD12 and REF. Dashed areas are statistically significant (see details in Sect. 2.4).

when calculated over the entire domain (-0.6 W/m^2). The effect of time-invariant aerosols on projected surface solar radiation is strong enough to altogether cancel the projected solar brightening in some regions (e.g., over Belgium) and even cause solar dimming over Scandinavia (not shown). These results are well aligned with the observations of [Gutiérrez et al., 2020], raising questions about the usability of the current EURO-CORDEX ensemble for future projections of solar power in Europe.

The substantial inconsistency in summer RSDS, which reaches -30 W/m^2 over Germany, is responsible for a weaker summer warming projected by ALD150AERO when compared to REF, with a mean difference of about -0.8 °C over Europe (computed over land points in the red box as in Fig. 2.a). The difference in projected temperature change can increase up to -1.5°C over Germany, or about 25% of the projected warming in REF for the same grid points.

Surprisingly, no significant differences are observed for total precipitation, evapotranspiration or cloud cover. The absence of significant effect on total cloud cover is partly caused by the fact that in our experiments, only the direct aerosols effect is considered. Therefore, the impact of aerosols representation may be even more important than shown here, when both direct and indirect aerosols effect are accounted for (not only direct radiative, but also the effect on cloud structure and lifetime).

3.5 Role of plant physiological response to CO2 variations

In the ALD150CO2 simulation, we reproduced the same experimental protocol as in the study of [Schwingshackl et al., 2019] for the COSMO_NOPHYS (no plant physiological response to CO2, see details in Sect. 2.3). By fixing the CO2 levels constant for the plants to the historical levels, the stomatal conductance of plants

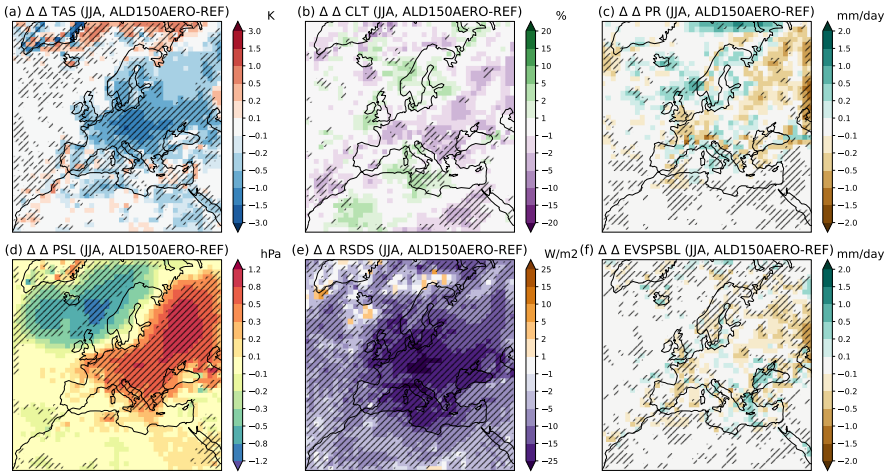

Fig. 10: Mean summer discrepancies in projected climate change signal between ALD150AERO and REF. Dashed areas are statistically significant (see details in Sect. 2.4).

in ALD150CO2 is higher throughout the RCP8.5 future simulation (2070-2099) than for REF. This causes a lower plant water use efficiency in ALD150CO2 and higher evapotranspiration levels (see Fig. 11.f).

In the case of REF/ALD150CO2 pair, the effects on evapotranspiration changes are mostly seen over Northern Europe. As a consequence of increased latent heat, ALD150CO2 also projects significantly weaker warming than REF for Northern Europe (about -0.5°C).

However, the effect is very weak in Southern and Central Europe, likely because of limited water availability in REF and the projected intensification of summer drying (see Fig. 4). This result is unexpected, as for example [Schwingshackl et al., 2019] shows that the strongest discrepancies in projected evapotranspiration change due to plant physiology effect occur in a band that spans from southern France to the Black Sea.

In any case, our results confirms that the physiological impact of CO2 on plants, and consequently on climate change, may be highly model-dependent [Boé, 2021].

3.6 Effect of GCM/RCM inconsistencies in atmospheric physics

Changing the atmospheric physics of the RCM model induces significant discrepancies in projected climate change for all variables and over a large part of the domain (Fig. 12). For example, a prominent new difference pattern is observed over the Arctic region, increasing the projected total cloud cover by the ALD150PHYS by up to 25%, reversing the projected decrease observed from REF over Scandinavia (Fig. 4.b). The changes in cloud cover are accompanied by changes in surface solar radiation, which can go as low as -38 W/m^2. This could be a consequence of a changed surface-atmosphere interaction resulting from the different atmospheric physics packages.

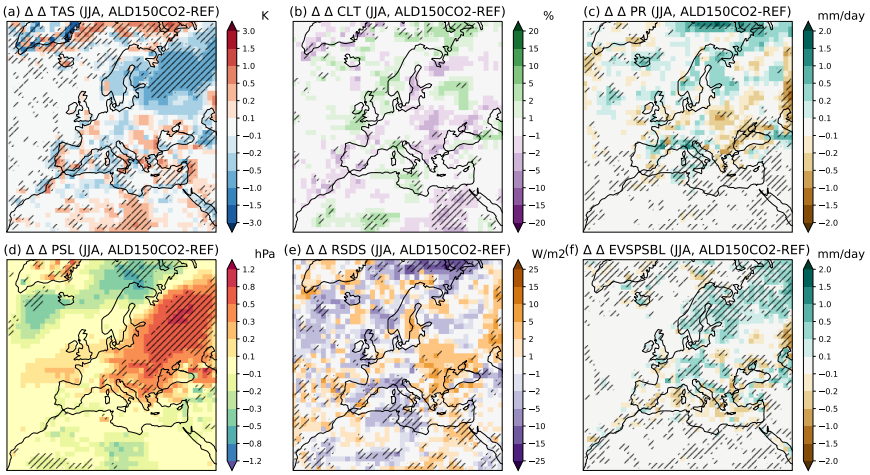

Fig. 11: Mean summer discrepancies in projected climate change signal between ALD150CO2 and REF. Dashed areas are statistically significant (see details in Sect. 2.4).

While the mean discrepancy in surface temperature is close to 0 over continental Europe, large discrepancies are observed regionally. For example, a weaker warming is observed in ALD150PHYS around the Mediterranean Sea (around -2°C over some points in the Iberian Peninsula) and another more robust warming pattern over Poland and the Baltic countries (around 1.2°C in some regions). These discrepancies structures seem to be driven by changes in cloud cover and the hydrological cycle through increased precipitation and evapotranspiration.

The increase in precipitation over continental Europe is about +0.6 mm/day, enough to cancel the projected drying by REF in most regions (see Fig. 4.c), and even reverse the signal over the Alps, causing a slight increase in humid conditions (not shown). It should be mentioned that the ALD150PHYS experiment uses the same radiative and land-surface schemes as the REF simulation (see Sect. 2.3). Therefore, the important inconsistencies in RSDS are likely a consequence of changes in total cloud cover and their vertical distribution, all driven by differences in the cloud, deep convection, and turbulence schemes between REF and ALD150PHYS.

Note that disentangling the exact origin of the difference between both physics is outside the scope of the current study. It would require testing step-by-step the different schemes from the CMIP6 physical package to the CMIP5 one. Instead, the REF/ALD150PHYS pair was designed to show that differences in physical parameterization may help reproduce the large scale inconsistencies which were previously documented for the EURO-CORDEX ensemble, therefore raising the question about the necessity of physically consistent GCM/RCM pairs.

The inconsistencies generated in ALD150PHYS should not be seen as a general rule, but simply due to the specifically used configuration. For a different pair of GCM/RCM, their unique set of physics schemes could lead to a different result. For example, a hindcast study evaluating the biases between ERA-Interim and a WRF multi-physics ensemble have shown that using different configurations

in microphysics, convection and radiation may lead to the development of unique bias patterns [Katragkou et al., 2015]. In another study, [Jerez et al., 2013], it was shown that choices in the parameterization schemes can cause a spread for the future projections similar in magnitude to the mean projected changes for the given scenario, and analogous to the spread obtained in a multi-model ensemble.

While previous multi-physics studies focused their attention on understanding the uncertainty in the choice of physical parameterization, we argue here that having inconsistent parameterization between the RCM and driving GCM causes important discrepancies in the large scale climate change signal, without necessarily adding more value. This may raise the question if it is methodologically correct to dynamically downscale GCMs, which are structurally very different from the RCM.

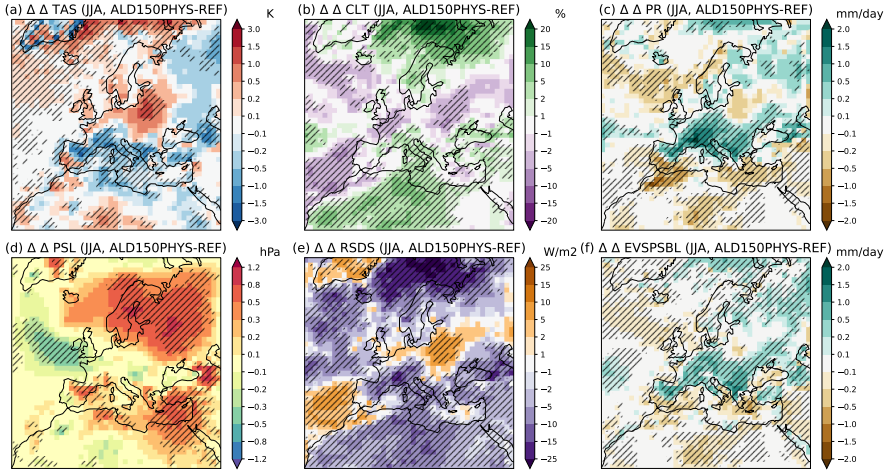

Fig. 12: Mean summer discrepancies in projected climate change signal between ALD150PHYS and REF. Dashed areas are statistically significant (see details in Sect. 2.4).

3.7 Combined effect of resolution, time-invariant aerosols, plant physiology and inconsistent atmospheric physics

Combining the set-up modifications from all previous sensitivity experiments, the ALD12MAX simulation shows robust differences in projected climate from the driving REF across all variables (see Fig. 13). The observed difference patterns resemble the ones between the EURO-CORDEX ensemble and the driving GCM ensemble [Boé et al., 2020], with much lower projected warming over Europe (about -1.3°C), weaker projected drying in Southern and Central Europe, and significant differences in the solar radiation at the surface, causing a solar dimming in ALD12MAX when compared to REF projected brightening (see Fig. 4).

The results show that the effects of different contributions are indeed additive, but not always linearly. For example, in terms of general trends, the ALD12MAX

seems to be a superposition of previous experiments, with even weaker warming over most of Europe, reduced drying conditions in Southern and Central Europe, solar dimming instead of brightening over most of the domain, and higher pressure at sea level over Europe. At the same time, the reduced surface energy availability mainly caused by the time-invariant aerosols, completely canceled the intensified evapotranspiration due to the plant physiological response to CO2 in Northern Europe (ALD150CO2 experiment). Such compensation mechanisms may exist in other GCM/RCM pairs in the EURO-CORDEX ensemble, which may show a relatively high degree of coherence for a particular variable, despite important structural differences. Taking this into account, it seems essential to evaluate GCM/RCM pairs using an extended set of variables (as in this study) to have a complete view of the observed inconsistencies or their absence.

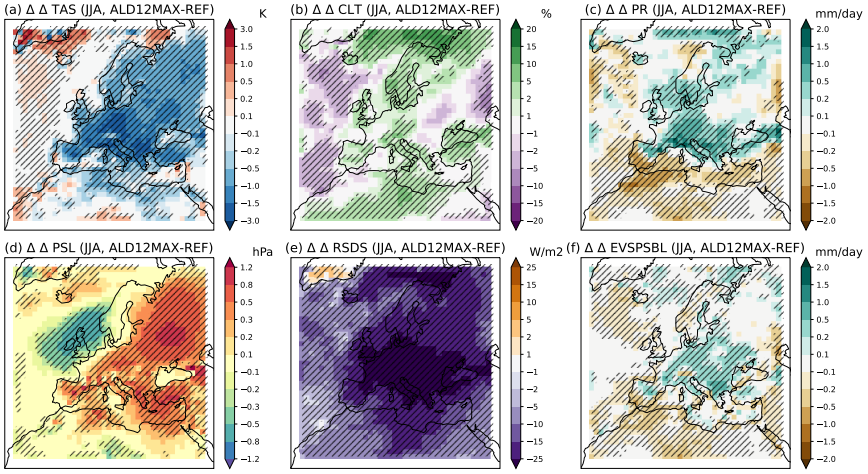

Fig. 13: Mean summer discrepancies in projected climate change signal between ALD12MAX and REF. Dashed areas are statistically significant (see details in Sect. 2.4).

3.8 Mean discrepancies over continental Europe

The previous sections presented how differences in resolution, external forcings, plant physiology, and atmospheric physics could generate large-scale discrepancies in surface variables such as temperature and precipitation between a RCM and its driving GCM that share the same dynamical core. We now want to place our experiments in the context of the 12km-resolution RCP8.5 EURO-CORDEX ensemble. To do so, the mean discrepancies were computed over continental Europe (red box in Fig. 2) for all 48 GCM/RCM pairs in the EURO-CORDEX ensemble and each experiment (see Fig. 14).

The most coherent pair in our experimental framework is the benchmark simulation REF/ALD150, which shares the same physics, external forcings, and almost

the same resolution. Except for precipitation, the mean discrepancies are close to 0 across all variables.

When analyzed only over continental Europe, the REF/ALD150CO2 seems even more coherent than REF/ALD150 (Fig. 14). This is because in the case of ALD150CO2, the effect of plant physiological response is seen only over Northern Europe, and the generated increase in evapotranspiration also corrected the slightly drier conditions of ALD150 over Central Europe (compare Fig. 8.c and 11.c).

For the REF/ALD12 pair, the mean discrepancies are close to 0 for most variables, except for the total precipitation and evapotranspiration (Fig. 14). These discrepancies in the hydrological cycle are mainly generated by the upscaled added value over the Alps (Fig. 9). Therefore, we conclude that, alone, the resolution effect cannot explain the large differences in surface solar radiation or temperature changes between EURO-CORDEX RCMs and their forcing GCMs [Boé et al., 2020].

The REF/ALD150AERO pair shows that the differences in aerosols representation between the GCM/RCM is likely a dominant factor in explaining the EURO-CORDEX surface solar radiation discrepancies in projected change. The differences in available solar energy at the surface translate into surface temperature discrepancies. As a result, the aerosol effect cause a weaker projected summer warming of the RCM vs the GCM by about 0.8 K over continental Europe (Fig. 14 TAS). This confirms the critical role of evolving aerosol forcing to explain GCM/RCM discrepancies in EURO-CORDEX and likely over other domains ([Gutiérrez et al., 2020, Boé et al., 2020, Chen, 2021]).

For the REF/ALD150PHYS pair, substantial inconsistencies are registered over continental Europe for precipitation and evapotranspiration. In magnitude, these discrepancies are comparable to the most inconsistent pairs of the EURO-CORDEX ensemble. For the other variables such as cloud cover and associated surface solar radiation, the spatially-averaged discrepancies are very small, but this is simply a consequence of canceling differences when aggregated over continental Europe (see Fig. 12). Even with only one run testing an inconsistent physics, whereas many other options would have been possible, we show the strong potential that GCM/RCM physics inconsistencies have to generate large-scale discrepancies of the surface variables, especially for the water cycle.

For most of the variables, the REF/ALD12MAX pair is the most (CLT, TAS, RSDS) or almost the most (PR, EVSPSBL) incoherent of the generated pairs, confirming that the inconsistency sources identified in our study are at least partly additive. In particular, it is worth mentioning that the REF/ALD12MAX pair shows the same level of inconsistency as the most incoherent pairs in EURO-CORDEX for RSDS. However, it is far from reproducing the outliers of the EURO-CORDEX ensemble, such as the -4.5K for surface warming or +2.0 mm/day in terms of precipitation changes (see outliers for TAS and PR in Fig. 14). The existence of such outliers may indicate that we did not identify yet all the sources of GCM/RCM inconsistency or that some pairs are truly incompatible, the RCM being unable for structural reason to follow the GCM imposed climate change at large-scale. In this context, the Tables SM. 6 and SM. 8 shows that the level of discrepancies may depend on the driving GCM or driven RCM, for example, with most RCMs struggling to maintain consistency with the MOHC-HadREM3 GCM.

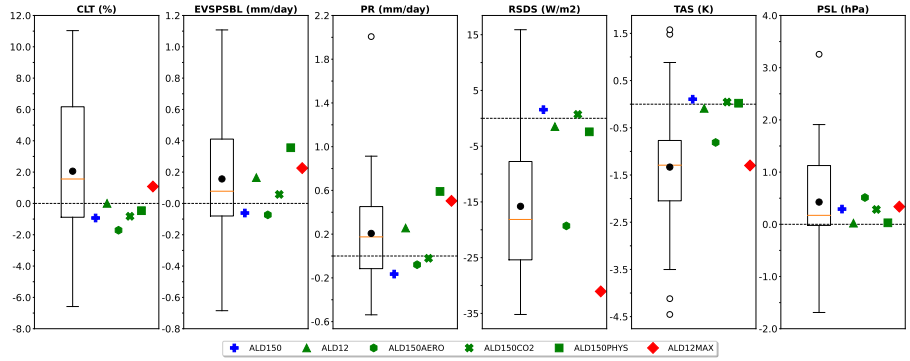

Fig. 14: Mean summer discrepancies in projected climate change over continental Europe, computed over the red box (see Fig. 2) for all EURO-CORDEX GCM/RCM pairs and all experiments presented in the study (detailed methods in Sect. 2.6). The whisker box represents the range of discrepancies in the EURO-CORDEX ensemble, with the black dot corresponding to the mean, the orange line to the median, the box covers the interquartile range (IQR) between Q1 and Q3 (25% and 75%), the whiskers covers the interval of Q1-1.5IQR and Q3+1.5IQR, and the outliers are represented as transparent circles. The mean discrepancies corresponding to each experiment are given using a set of colored markers, with a legend provided under the plot.

3.9 Fraction of incoherent grid points

Computing the mean discrepancy over large regions is a useful and often employed method to gain a general understanding of differences patterns. At the same time, it has two important limitations. First, it may show important signal, when there are almost no statistically significant points (e.g. in Fig. 14 for the CLT variable of ALD150AERO experiment). Secondly, it may show very weak signal if the difference patterns have high heterogeneity over the region of integration (e.g. in Fig. 14 for the TAS variable for the ALD150PHYS experiment).

In order to solve these limitations, we propose the fraction of incoherent grid points, as a simple yet very effective way to quantify the effect of each mechanism on the overall coherence of the GCM/RCM pair (see details Sect. 2.7). Using this metric, the REF/ALD150 remains the most coherent pair, followed by ALD150CO2, ALD12, ALD150AERO, ALD150PHYS, and finally ALD12MAX (see mean incoherent fractions in Table 2). In terms of mean inconsistencies, the benchmark pair REF/ALD150 was the most coherent pair by far, with a mean summer incoherent fraction of 16.4% compared to REF/ALD12 with 27.4%, or the EURO-CORDEX mean of 61.7% (see Table SM. 6).

Judging by the mean discrepancies alone (see Fig. 14), it can be challenging to measure the level of inconsistency due to canceling positive and negative differences. In this regard, Table 2 suggest that the differences in aerosols and the physical representation of atmospheric processes are probably the two main factors in generating the large-scale discrepancies in the EURO-CORDEX ensemble.

Model	TAS	RSDS	PR	EVSPSBL	CLT	PSL	Mean
ALD150	10.5	12.0	26.6	35.1	7.4	6.8	16.4
ALD12	22.3	29.3	34.3	44.1	24.2	10.3	27.4
ALD150AERO	39.5	96.9	23.2	35.2	10.1	41.0	41.0
ALD150CO2	19.0	7.2	17.0	39.5	5.2	21.0	18.1
ALD150PHYS	20.8	58.3	41.5	57.7	46.7	41.0	44.3
ALD12MAX	76.2	100.0	52.2	60.5	45.2	61.6	66.0

Value
[0,20)
[20,40)
[40,60)
[60,80)
[80,100]

Table 2: Fraction of incoherent grid points computed over the land part of the standard EUR-11 domain (black box in Fig. 2) for each experiment (see methods 2.7) using the discrepancies in projected summer climate change. The fraction is calculated as the ratio between the number of land points where the differences in projected change for a given variable between REF and ALD are statistically significant and the total number of land points in the EUR-11 domain. The mean across all six variables is also provided to measure the overall inconsistency. The color code indicated aside the Table is subjective and targets to propose a synthetic vision of the results.

Concerning the discrepancies in cloud cover between the GCM/RCM ensembles, the ALD150PHYS vs. ALD12 suggests that the differences in CLT are more likely generated by structural differences in the physical parameterization of the GCM/RCM, rather than the upscaled added value, something to be explored more in the future studies.

When placed in the context of the EURO-CORDEX ensemble (see Fig. 15), the REF/ALD150 pair achieve a much higher overall level of consistency than any other GCM/RCM pair from the 48 available pairs in the ensemble (see Table SM. 2-9 for a detailed overview of inconsistencies in the EURO-CORDEX for each GCM/RCM pair). At the same time, the REF/ALD12 pair shows that the effect of upscaled added value alone cannot justify the discrepancies seen in the ensemble, primarily affecting the hydrological cycle in areas with complex surfaces (see Fig. 9). From these results, it seems that the role of plant physiology is insignificant in explaining the inconsistencies seen in the EURO-CORDEX. At the same time, it should be noted that the effect of plant physiology may be highly model dependent [Boé, 2021]. The time-invariant aerosols effect taken into account in REF/ALD150AERO seems to reproduce the inconsistencies in projected surface solar radiation, but not enough to explain inconsistencies in other variables. Therefore, it is suggested that the differences in temperature and other variables are a consequence of the combined effect of physics, aerosols, resolution, and plant physiology (ordered by likely intensity of contribution). Even if the REF/ALD12MAX shows a high level of inconsistency, it is still below the outliers of the EURO-CORDEX ensemble (except for RSDS). This probably means that our modelling framework based on only one GCM/RCM pair and with a limited number of tests in particular for the physics is not able to uncover all sources of inconsistency in EURO-CORDEX.

It should be noted that these suggestions are based on experiments done with a single GCM/RCM pair. In order to obtain more decisive evidence, additional

experiments may be required, including with other GCM/RCM pairs. We are also aware that the modelling framework proposed here may have missed key sources of GCM/RCM inconsistency in the EURO-CORDEX ensemble such as air-sea coupling, aerosols-cloud interaction and other external forcings like land-use change, volcanic aerosols, GHG or ozone. For example, the air-sea coupling was previously shown to be relevant in the representation of heat exchanges between ocean and atmosphere, especially in the Baltic region, affecting temperature extremes near coastline and offshore winds intensity [Gröger et al., 2021]. A different study, [Jerez et al., 2018], showed that some regional climate models did not include evolving GHG as forcing which under 1.5°C global warming translates into a non-negligible impact of up to 1°C in the regional projections of surface temperature. Finally, a more extensive list of references on the neglected Earth processes in regional climate modelling and their impact can be found in the dedicated section, 10.3.1.2, of the IPCC AR6 Chapter 10 [Doblas-Reyes et al., 2021].

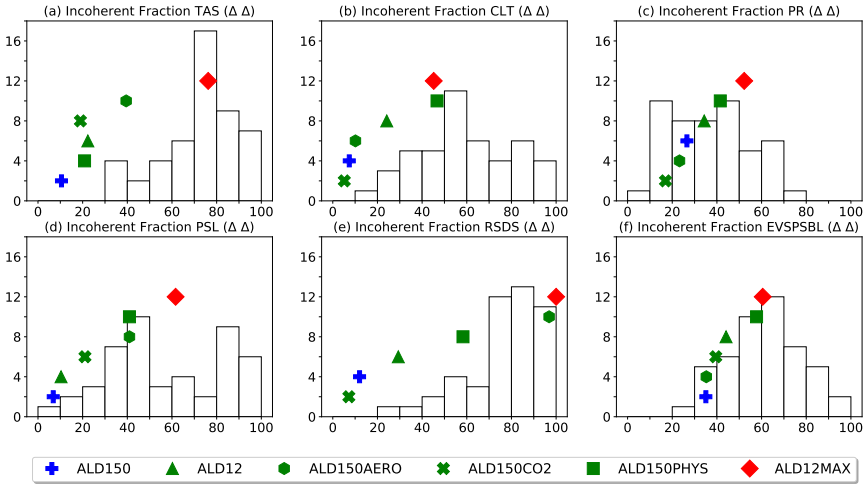

Fig. 15: Percentage of incoherent grid points computed over the land part of the standard EUR-11 domain (black box in Fig. 2) for each GCM/RCM pair in the EURO-CORDEX ensemble and each experiment (see methods 2.7) using the discrepancies in projected summer climate change. The fraction is calculated as the ratio between the number of land points where the differences in projected change for a given variable between REF and ALD are statistically significant and the total number of land points in the EUR-11 domain. The histograms show the distribution of incoherent fractions in the EURO-CORDEX 48 GCM/RCM pairs, and the fractions for experiments are represented using colored markers with provided legend below the figure.

3.10 Impact of each mechanism on the GCM/RCM large scale dynamics

In previous sections, we quantified the role of various factors in explaining the large-scale discrepancies seen in a typical GCM/RCM pair over Europe in projected climate change. At the same time, little is known about how such differences in resolution, physics, or forcing can affect the ability of the GCM to successfully impose its large scale dynamics onto the driven RCM.

Even if there is not yet a consensus in the RCM community, let's assume for this section, that the large scale circulation should be mostly unaffected within the RCM domain, therefore following the Tenet 5.a in [Laprise et al., 2008]. Significant deviations from this rule are usually associated with the internal variability or some structural differences between the RCM and driving GCM. To understand how Tenet 5.a holds in the context of our study, a von Storch analysis was performed [von Storch et al., 2000]. Such an analysis allows quantifying the similarity in the kinetic energy variance encoded in the large atmospheric modes ($\lambda \in [1000, 6000]$ km) of the REF and driven ALD (see Sect. 2.8 for details). In Fig. 16 we summarize the results of this analysis for each experiment and both HIST (1970-1999) and RCP8.5 (2070-2099). The results are given here for summer, but a similar analysis was done for winter and provided in Supplementary materials (see Fig. SM. 14).

In terms of mean inconsistencies, the benchmark pair REF/ALD150 was the most coherent pair by far when compared with other experiments or EURO-CORDEX pairs (see Table 2 and SM. 6). It is therefore surprising to find that in terms of the large-scale dynamics coherence, the ALD12 simulation is better than ALD150, with a median proportion of summer represented variance for the Historical experiment of 0.84 vs. 0.80 and a smaller interquartile range of 0.17 vs. 0.25 respectively (see Fig. 16). There are three possible explanations for this. Firstly, the domain size of the ALD12 configuration is slightly smaller than the ALD150 one (about 300 km from each side, see Fig. 2), which likely implies a higher internal variability in ALD150 [Alexandru et al., 2007]. Secondly, it was suggested that the forcing from the driving field on the RCM simulations is weaker over the Arctic than the mid-latitude [Rinke and Dethloff, 2000]. The fact that the ALD150 North boundary is further into the Arctic could therefore contribute to a lower overall large-scale circulation coherence. Finally, the ALD150 configuration is abnormal for an RCM as it has a very low resolution, and choosing the number of grid points within the relaxation zone [Davies, 1976] is not straightforward. We choose here 3 points (about 450 km), but that may be too small in the number of grid points or too large in kilometers to ensure a good representation of the large-scale dynamics at high frequency. Further tests are therefore required.

Despite significant discrepancies in projected changes in temperature and pressure at sea level (see Fig. 10), it was found that the ALD150AERO large-scale dynamic has only slightly deteriorated when compared to the ALD150 benchmark experiment for the same period RCP8.5 (2070-2099, see Fig. 16). This proves that the differences in external forcing representation between GCM/RCM pairs are not a major concern for their large-scale circulation coherence. The same conclusion applies to the REF/ALD150CO2 pair, with the negative impact of plant physiology on large-scale dynamics coherence being only slightly stronger than the aerosols one.

The conclusion is the opposite for the REF/ALD150PHYS pair, with its lowest median proportion of represented variance for the HIST period, of about 0.74

and largest interquartile range of about 0.34. This experience confirms, but this time from a purely dynamical point of view, that the differences in the model's physical schemes are likely to strongly modulate the irreproducible component of the climate variability [Crétat and Pohl, 2012].

Recently, multi-physics ensembles have been used as a tool to explore the physical realism of a model [Katragkou et al., 2015, García-Díez et al., 2015, Lavin-Gullon et al., 2021]. However, we argue that the interpretation of such ensembles should be made with precaution because the observed biases between the RCM and observations generated by different physical parameterization are likely not only a consequence of changes in physics but also of the increased or decreased capacity of the RCM to follow the GCM large-scale circulation.

Despite being the most inconsistent pair in terms of mean discrepancies over continental Europe and fraction of incoherent land grid points over the EUR-11 domain for all variables, the ALD12MAX experiment is second best in terms of large-scale dynamics coherence. For example, for the historical period and summer season, the ALD12MAX has a median of represented variance of about 0.82 and an interquartile range of about 0.21, compared to 0.80 and 0.25 for the ALD150 experiment respectively. This result is difficult to explain at the moment, as for example, for winter, the opposite is observed, with ALD12MAX being the second worst, after ALD150PHYS (see Fig. SM. 14).

When comparing the results for the HIST and RCP8.5 experiments, it was found that the capacity of the RCM to follow the GCM large-scale circulation increased under future atmospheric conditions for all the experiments. This could be a consequence of changes in future synoptic conditions, which are less favorable to internal variability ([Lucas-Picher et al., 2008, Sanchez-Gomez et al., 2009, Sieck et al., 2016]). However, more investigations are required to say if this result can be generalized to any GCM/RCM pair. In particular, a large ensemble of simulations, as in [Sanchez-Gomez and Somot, 2018] but performed separately for the historical and the future periods, thus allowing to compare the level of RCM internal variability in both periods, would be of interest.

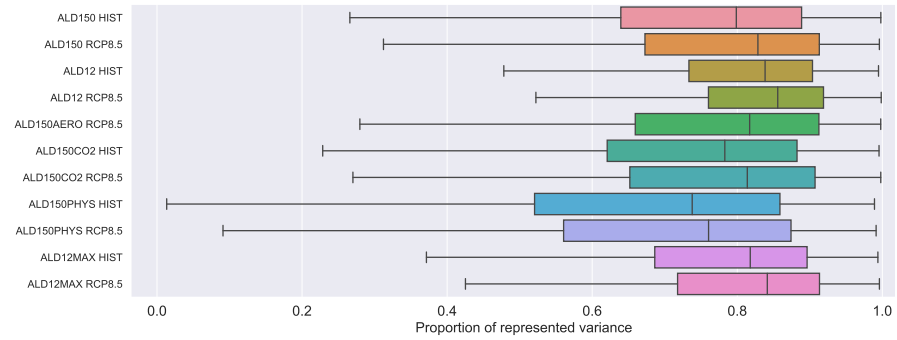

Fig. 16: Box-and-whiskers plot representation of the proportion of represented variance, $P(t)$, computed over the HIST (1970-1999) and RCP8.5 (2070-2099) periods for each experiment for summer. The boxes are delimited by the 25th and 75th percentiles (Q1 and Q3), with the median in between. The range between Q1 and Q3 is the interquartile range (IQR). The minimum and maximum whiskers are given by Q1 - 1.5IQR and Q3+1.5IQR. A value of $P(t)$ close to 1 means that the REF and ALD agree on the large-scale circulation. A $P(t)$ smaller than 1 means that the REF and ALD large-scale circulation differ.

4 Conclusion

The EURO-CORDEX RCMs project at the end of the 21st century over a large area of Europe, a summer warming 1.5–2 K colder, and a much smaller decrease in precipitation of 5%, versus 20% in their driving GCMs for the high-emission scenario RCP8.5 [Boé et al., 2020]. Considering the importance of reliable regional climate information for vulnerability, impact, and adaptation applications, it is essential to understand the causes of these inconsistencies at large-scale. This study focuses mainly on the mechanisms behind large-scale summer inconsistencies between GCM/RCM climate change projections between the historical period (1970-1999) and the end of the century in the RCP8.5 high-emission scenario (2070-2099). At the same time, all the results are also provided for winter in the Supplementary materials.

In order to explore this inconsistency, a unique experimental protocol was implemented for the EURO-CORDEX domain based on a perfect-model approach in the same spirit as the popular Big-Brother/Little-Brother experiment in the RCM literature [Denis et al., 2002b], with the GCM/RCM pair sharing the same dynamical core and physical parameterizations (version 6 of the CNRM climate models used in CMIP6 and EURO-CORDEX), external forcings (sea surface temperature, sea ice cover, aerosols, and ozone), vertical discretization (91 levels) and horizontal resolution (about 150km). Furthermore, setting this perfect benchmark GCM/RCM pair was allowed by the seamless strategy chosen in the development of version 6 of the GCM CNRM-ARPEGE and the RCM CNRM-ALADIN.

Once the capacity to generate a coherent GCM/RCM pair is confirmed, we deviate from the perfect approach by introducing imperfections. By isolating each effect, we can quantify the role of each mechanism in generating large-scale discrepancies. The mechanisms investigated in this article are the effect of the resolution,

time-invariant aerosols in future projections, plant physiology, and atmospheric physics.

A detailed analysis of the historical period for the benchmark experiment shows that a relatively good coherence is achievable, especially for surface temperature and precipitation, with the mean discrepancies over the entire RCM domain of -0.1 °C for temperature and +0.1 mm/day for precipitations. At the same time, it can be challenging to achieve a high degree of coherence on each grid point independently. This happens because a typical GCM/RCM pair operates on different native grids. Consequently, the topography, the land-sea mask, the sea ice limit, and probably other geometry-dependent elements are defined slightly differently in the GCM vs. the RCM. In particular, it was found that the difference in topography between the GCM and the RCM in the perfect pair explains 70% of differences in surface temperature and 25% in cloud cover over the regions with high relief for the historical period (1970-1999). The role of sea-borders definition being more modest when compared to the topography, explaining about 20% of discrepancies in the total cloud cover and 10% in evaporation over the grid points qualified as the coastline.

Despite the identified modeling limitations, it was found that their contributions to the large-scale discrepancies are systematic across all variables and cancel out when the climate change signal is computed. Over the six studied variables (surface temperature, surface solar radiation, total cloud cover, precipitation, and evapotranspiration), on average only 16% of the EUR-11 domain land points had a statistically significant difference. This is an extremely low level of inconsistency when compared to the 62% average of the 12.5 km RCP8.5 EURO-CORDEX ensemble and shows the validity of our approach. Being able to achieve this almost perfect GCM/RCM pair over the EURO-CORDEX domain opens the door to a wide range of studies to better understand the GCM/RCM discrepancies.

The next step after confirming the consistency of the seamless GCM/RCM pair is to deviate from this perfect approach by introducing "imperfections".

By increasing the horizontal resolution of the RCM to the standard EUR-11 grid at 12 km, we find that the upscaled added value generate discrepancies in the projected climate change for the water cycle, including total precipitation and cloud cover. At the same time, these differences are mostly located in the areas with complex terrain, while other parts of the domain maintain consistency. The effect of time-invariant aerosols is much stronger, with inconsistencies in projected summer surface solar radiation attaining a maximum of -30W/m^2 over Germany, and reducing the overall RCM summer warming by about -0.8°C over Europe. Our results for the aerosol test confirm previous literature [Gutiérrez et al., 2020, Boé et al., 2020]. In our experiments, the plant physiological response to increasing atmospheric CO2 concentrations is identified as an additional source of discrepancy. If this effect is present in the GCM, but not in the RCM, we find a weaker projected summer warming of the RCM by about -0.5°C over Northern Europe. These differences in surface temperature change are driven by higher evapotranspiration rates in the RCM vs GCM. We find that the plant physiology effect is seen mostly in Northern Europe, but this may be highly model dependent. Our results for the plant effect partially confirms the finding of [Schwingshackl et al., 2019]. By introducing changes to the convection, turbulence and cloud physics schemes of the RCM, we find that differences in the physical parameterization may be one of the most important factors in the generation of GCM/RCM inconsistencies.

The variables which were the most sensitive to this test are related to the hydrological cycle, notably +0.6mm/day for total precipitation and +0.4mm/day for total evapotranspiration over the continental Europe. The final experiment was to combine all the previously mentioned mechanisms. By doing so, we find that the contribution of each mechanism is mainly additive, leading to a highly inconsistent GCM/RCM pair in terms of the projected summer climate change. The obtained results for this experiment are very similar to the discrepancies patterns observed for the EURO-CORDEX ensemble, with about 66% of land points over the EUR-11 domain having statistically significant differences in projected climate signal across all analyzed variables.

Additionally, we studied how the structural differences between the GCM/RCM pairs may also affect the consistency of the large-scale dynamics on a day-to-day basis, resulting in larger internal variability in RCMs. We find that the GCM/RCM resolution gap, differences in the aerosols forcing, or plant physiology are not damaging much the ability of the GCM to impose its large-scale circulation onto the RCM. On the contrary, significant differences in the atmosphere's physical parameterization, as well as domain size and position, can significantly damage the GCM/RCM large-scale circulation coherence. Therefore, we conclude that the dynamical inconsistencies previously detected by the RCM community, known under the name of internal variability, are partially generated by the structural differences between the GCM/RCM pairs. While not shown in this study, when inconsistencies in physical parameterization are unavoidable between the GCM/RCM pair, the large-scale dynamical consistency may be improved by using large-scale nudging techniques [von Storch et al., 2000].

The main takeaway message from this study is that the upscaled added value of regional climate models is not sufficient to explain the large-scale discrepancies in projected climate change observed for the 12.5 km EURO-CORDEX ensemble. Instead, the main factors driving these discrepancies are the differences in external forcing and in physical parameterization. While the role of physics in GCM/RCM inconsistencies is yet to be untangled by future studies, it is clear that efforts should be made by the RCM community in implementing the time-evolving aerosols in their future simulations.

The fact that important differences in the GCM/RCM physical parameterization will inevitably generate inconsistencies in the large-scale climate change projections and damage the day-to-day consistency of the large-scale circulation raises a philosophical discussion about the validity of the current dynamical downscaling practice in which the GCM/RCM physical consistency is not a criterion in the GCM selection. So what are the possible solutions?

A short-term solution could be to reduce the CORDEX GCM/RCM matrix by preferentially running the "structurally more compatible" GCM/RCM pairs. In order to establish such compatibility, new large-scale coherence-based metrics should be developed, such as the one proposed in our study, which is the fraction of incoherent land points over the EUR-11 domain. In this context, the multi-physics ensembles should be used with precaution because of the associated risks to taking us even further from the validity of the Tenet 5.a axiom by damaging the GCM/RCM large scale dynamical coherence [Laprise et al., 2008].

In addition to using the classical nested-grid approach, the regional climate modeling community may benefit from relying more on stretched grid models. The important advantages of variable resolution stretched-grid GCMs are that they do

not require any lateral boundary conditions/forcing and are free of the associated undesirable computational problems; as a result, they provide self-consistent interactions between global and regional scales of motion and their associated phenomena as in uniform grid GCMs [Fox-Rabinovitz et al., 2006].

In the long term, the community may also benefit from more studies on the GCM/RCM inconsistencies in projected climate change. For example, we did not cover in our study the potential role of ocean coupling or land-use change. We are also limited by testing only one GCM/RCM pair. Such studies will eventually help us identify the next steps in improving regional climate modelling and design more strict downscaling protocols to assure the community with more reliable high resolution climate information.

Declarations

Ethical Approval Not applicable.

Competing Interests The authors have no relevant financial or non-financial interests to disclose.

Authors' contributions Ioan Sabin Taranu wrote the main manuscript text, participated in the process of conceptualization and methodology design, did an important part of the formal analysis and visualization. Samuel Somot designed and supervised the project. Antoinette Alias participated in the process of design, running and validation of the experiments. Julien Boé did the analysis of the EURO-CORDEX ensemble data. Christine Delire contributed to the design and validation of the experiment on the plant physiology effect. All authors actively followed the project with regular exchanges, reviewed the manuscript and provided critical feedback.

Funding The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Availability of data and materials The datasets and the codes generated for the current study are available under request from the corresponding author.

References

[Alexandru et al., 2007] Alexandru, A., de Elia, R., and Laprise, R. (2007). Internal variability in regional climate downscaling at the seasonal scale. *Monthly Weather Review*, 135(9):3221–3238.

[Arias et al., 2021] Arias, P. A., Bellouin, N., Coppola, E., Jones, R. G., Krinner, G., Marotzke, J., Naik, V., Palmer, M. D., Plattner, G.-K., Rogelj, J., Rojas, M., Sillmann, J., Storelvmo, T., Thorne, P. W., Trewin, B., Rao, K. A., Adhikary, B., Allan, R. P., Armour, K., Bala, G., Barimalala, R., Berger, S., Canadell, J. G., Cassou, C., Cherchi, A., Collins, W., Collins, W. D., Connors, S. L., Corti, S., Cruz, F., Dentener, F. J., Dereczynski, C., Luca, A. D., Niang, A. D., Doblas-Reyes, F. J., Dosio, A., Douville, H., Engelbrecht, F., Eyring, V., Fischer, E., Forster, P., Fox-Kemper, B., Fuglestvedt, J. S., Fyfe, J. C., Gillett, N. P., Goldfarb, L., Gorodetskaya, I., Gutierrez, J. M., Hamdi, R., Hawkins, E., Hewitt, H. T., Hope, P., Islam, A. S., Jones, C., Kaufman, D. S., Kopp, R. E., Kosaka, Y., Kossin, J., Krakovska, S., Lee, J.-Y., Li, J., Mauritsen, T., Maycock, T. K., Meinshausen, M., Min, S.-K., Monteiro, P. M. S., Ngo-Duc, T., Otto, F., Pinto, I., Pirani, A., Raghavan, K., Ranasinghe, R., Ruane, A. C., Ruiz, L., Sallée, J.-B., Samset, B. H., Sathyendranath, S., Seneviratne, S. I., Sörensson, A. A., Szopa, S., Takayabu, I., Treguier, A.-M., van den Hurk, B., Vautard, R., von Schuckmann, K., Zaehle, S., Zhang, X., Zickfeld, K., Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S. L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M. I., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J. B. R., Maycock, T. K., Waterfield, T., Yelekçi, O., Yu, R., and Zhou, B. (2021). Technical summary. in: Climate change 2021: The physical science basis. contribution of working group i to the sixth assessment report of the intergovernmental panel on climate change.

[Bador et al., 2017] Bador, M., Terray, L., Boe, J., Somot, S., Alias, A., Gibelin, A.-L., and Dubuisson, B. (2017). Future summer mega-heatwave and record-breaking temperatures in a warmer france climate. *Environmental Research Letters*, 12(7):074025.

[Bartók et al., 2017] Bartók, B., Wild, M., Folini, D., Lüthi, D., Kotlarski, S., Schär, C., Vautard, R., Jerez, S., and Imecs, Z. (2017). Projected changes in surface solar radiation in cmip5 global climate models and in euro-cordex regional climate models for europe. *Climate dynamics*, 49(7):2665–2683.

[Bellprat et al., 2016] Bellprat, O., Kotlarski, S., Lüthi, D., De Elía, R., Frigon, A., Laprise, R., and Schär, C. (2016). Objective calibration of regional climate models: application over europe and north america. *Journal of Climate*, 29(2):819–838.

[Boé, 2021] Boé, J. (2021). The physiological effect of co2 on the hydrological cycle in summer over europe and land-atmosphere interactions. *Climatic Change*, 167(1):1–20.

[Boé et al., 2020] Boé, J., Somot, S., Corre, L., and Nabat, P. (2020). Large discrepancies in summer climate change over europe as projected by global and regional climate models: causes and consequences. *Climate Dynamics*, 54(5):2981–3002.

[Chen, 2021] Chen, L. (2021). Uncertainties in solar radiation assessment in the united states using climate models. *Climate Dynamics*, 56(1):665–678.

[Colin et al., 2010] Colin, J., Déqué, M., Radu, R., and Somot, S. (2010). Sensitivity study of heavy precipitation in limited area model climate simulations: influence of the size of the domain and the use of the spectral nudging technique. *Tellus A: Dynamic Meteorology and Oceanography*, 62(5):591–604.

[Coppola et al., 2021] Coppola, E., Nogherotto, R., Ciarlò, J. M., Giorgi, F., van Meijgaard, E., Kadygrov, N., et al. (2021). Assessment of the european climate projections as simulated by the large euro-cordex regional and global climate model ensemble. *Journal of Geophysical Research: Atmospheres*, 126(4).

[Crétat and Pohl, 2012] Crétat, J. and Pohl, B. (2012). How physical parameterizations can modulate internal variability in a regional climate model. *Journal of the atmospheric Sciences*, 69(2):714–724.

[Daniel et al., 2019] Daniel, M., Lemonsu, A., Déqué, M., Somot, S., Alias, A., and Masson, V. (2019). Benefits of explicit urban parameterization in regional climate modeling to study climate and city interactions. *Climate Dynamics*, 52(5):2745–2764.

[Davies, 1976] Davies, H. (1976). A lateral boundary formulation for multi-level prediction models. *Quarterly Journal of the Royal Meteorological Society*, 102(432):405–418.

[Decharme et al., 2019] Decharme, B., Delire, C., Minvielle, M., Colin, J., Vergnes, J.-P., Alias, A., Saint-Martin, D., Séférian, R., Sénési, S., and Voldoire, A. (2019). Recent changes in the isba-ctrip land surface system for use in the cnrm-cm6 climate model and in global off-line hydrological applications. *Journal of Advances in Modeling Earth Systems*, 11(5):1207–1252.

[Denis et al., 2002a] Denis, B., Cote, J., and Laprise, R. (2002a). Spectral decomposition of two-dimensional atmospheric fields on limited-area domains using the discrete cosine transform (dct). *Monthly Weather Review*, 130(7):1812–1829.

[Denis et al., 2002b] Denis, B., Laprise, R., Caya, D., and Côté, J. (2002b). Downscaling ability of one-way nested regional climate models: the big-brother experiment. *Climate Dynamics*, 18(8):627–646.

[Di Luca et al., 2016] Di Luca, A., Argüeso, D., Evans, J. P., de Elía, R., and Laprise, R. (2016). Quantifying the overall added value of dynamical downscaling and the contribution from different spatial scales. *Journal of Geophysical Research: Atmospheres*, 121(4):1575–1590.

[Doblas-Reyes et al., 2021] Doblas-Reyes, F., Sorensson, A., Almazroui, M., Dosio, A., Gutowski, W., Haarsma, R., Hamdi, R., Hewitson, B., Kwon, W.-T., Lamptey, B., et al. (2021). Linking global to regional climate change.

[Eyring et al., 2016] Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., and Taylor, K. E. (2016). Overview of the coupled model intercomparison project phase 6 (cmip6) experimental design and organization. *Geoscientific Model Development*, 9(5):1937–1958.

[Fox-Rabinovitz et al., 2006] Fox-Rabinovitz, M., Côté, J., Dugas, B., Déqué, M., and McGregor, J. L. (2006). Variable resolution general circulation models: Stretched-grid model intercomparison project (sgmip). *Journal of Geophysical Research: Atmospheres*, 111(D16).

[García-Díez et al., 2015] García-Díez, M., Fernández, J., and Vautard, R. (2015). An rcm multi-physics ensemble over europe: multi-variable evaluation to avoid error compensation. *Climate dynamics*, 45(11):3141–3156.

[Giorgi et al., 2009] Giorgi, F., Jones, C., Asrar, G. R., et al. (2009). Addressing climate information needs at the regional level: the cordex framework. *World Meteorological Organization (WMO) Bulletin*, 58(3):175.

[Giorgi and Mearns, 1999] Giorgi, F. and Mearns, L. O. (1999). Introduction to special section: Regional climate modeling revisited.

[Giorgi et al., 2016] Giorgi, F., Torma, C., Coppola, E., Ban, N., Schär, C., and Somot, S. (2016). Enhanced summer convective rainfall at alpine high elevations in response to climate warming. *Nature Geoscience*, 9(8):584–589.

[Gröger et al., 2021] Gröger, M., Dieterich, C., and Meier, H. (2021). Is interactive air sea coupling relevant for simulating the future climate of europe? *Climate Dynamics*, 56(1):491–514.

[Guérémy, 2011] Guérémy, J. (2011). A continuous buoyancy based convection scheme: one- and three-dimensional validation. *Tellus A: Dynamic Meteorology and Oceanography*, 63(4):687–706.

[Gutiérrez et al., 2020] Gutiérrez, C., Somot, S., Nabat, P., Mallet, M., Corre, L., Van Meijgaard, E., Perpiñán, O., and Gaertner, M. Á. (2020). Future evolution of surface solar radiation and photovoltaic potential in europe: investigating the role of aerosols. *Environmental Research Letters*, 15(3):034035.

[Gutiérrez et al., 2021] Gutiérrez, Jones, R. G., Narisma, G. T., Alves, L. M., Amjad, M., Gorodetskaya, I. V., Grose, M., Klutse, N. A. B., Krakovska, S., Li, J., Martínez-Castro, D., Mearns, L. O., Mernild, S. H., Ngo-Duc, T., van den Hurk, B., Yoon, J.-H., 2021, Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S. L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M. I., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J. B. R., Maycock, T. K., Waterfield, T., Yelekçi, O., Yu, R., and Zhou, B. (2021). Atlas. in climate change 2021: The physical science basis. contribution of working group i to the sixth assessment report of the intergovernmental panel on climate change.

[Hourdin et al., 2017] Hourdin, F., Mauritsen, T., Gettelman, A., Golaz, J.-C., Balaji, V., Duan, Q., Folini, D., Ji, D., Klocke, D., Qian, Y., et al. (2017). The art and science of climate model tuning. *Bulletin of the American Meteorological Society*, 98(3):589–602.

[Jerez et al., 2018] Jerez, S., López-Romero, J., Turco, M., Jiménez-Guerrero, P., Vautard, R., and Montávez, J. (2018). Impact of evolving greenhouse gas forcing on the warming signal in regional climate model experiments. *Nature communications*, 9(1):1–7.

[Jerez et al., 2013] Jerez, S., Montavez, J. P., Gomez-Navarro, J. J., Lorente-Plazas, R., Garcia-Valero, J. A., and Jimenez-Guerrero, P. (2013). A multi-physics ensemble of regional climate change projections over the iberian peninsula. *Climate dynamics*, 41(7):1749–1768.

[Katragkou et al., 2015] Katragkou, E., García-Díez, M., Vautard, R., Sobolowski, S., Zanis, P., Alexandri, G., Cardoso, R. M., Colette, A., Fernandez, J., Gobiet, A., et al. (2015). Regional climate hindcast simulations within euro-cordex: evaluation of a wrf multi-physics ensemble. *Geoscientific Model Development*, 8(3):603–618.

[Laprise et al., 2008] Laprise, R., De Elia, R., Caya, D., Biner, S., Lucas-Picher, P., Diaconescu, E., Leduc, M., Alexandru, A., and Separovic, L. (2008). Challenging some tenets of regional climate modelling. *Meteorology and Atmospheric Physics*, 100(1):3–22.

[Lavin-Gullon et al., 2021] Lavin-Gullon, A., Fernandez, J., Bastin, S., Cardoso, R. M., Fita, L., Giannaros, T. M., Goergen, K., Gutiérrez, J. M., Kartsios, S., Katragkou, E., et al. (2021). Internal variability versus multi-physics uncertainty in a regional climate model. *International Journal of Climatology*, 41:E656–E671.

[Li et al., 1995] Li, D., Shine, K., and Gray, L. (1995). The role of ozone-induced diabatic heating anomalies in the quasi-biennial oscillation. *Quarterly Journal of the Royal Meteorological Society*, 121(524):937–943.

[Lucas-Picher et al., 2008] Lucas-Picher, P., Caya, D., Biner, S., and Laprise, R. (2008). Quantification of the lateral boundary forcing of a regional climate model using an aging tracer. *Monthly Weather Review*, 136(12):4980–4996.

[Masson-Delmotte et al., 2021] Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S. L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M., et al. (2021). Climate change 2021: the physical science basis. *Contribution of working group I to the sixth assessment report of the intergovernmental panel on climate change*, page 2.

[Mélia, 2002] Mélia, D. S. (2002). A global coupled sea ice–ocean model. *Ocean Modelling*, 4(2):137–172.

[Miguez-Macho et al., 2004] Miguez-Macho, G., Stenchikov, G. L., and Robock, A. (2004). Spectral nudging to eliminate the effects of domain position and geometry in regional climate model simulations. *Journal of Geophysical Research: Atmospheres*, 109(D13).

[Nabat et al., 2020] Nabat, P., Somot, S., Cassou, C., Mallet, M., Michou, M., Bouniol, D., Decharme, B., Drugé, T., Roehrig, R., and Saint-Martin, D. (2020). Modulation of radiative aerosols effects by atmospheric circulation over the euro-mediterranean region. *Atmospheric Chemistry and Physics*, 20(14):8315–8349.

[Rinke and Dethloff, 2000] Rinke, A. and Dethloff, K. (2000). The influence of initial and boundary conditions on the climate of the arctic in a regional climate model. *Climate Res*, 14:101–113.

[Roehrig et al., 2020] Roehrig, R., Beau, I., Saint-Martin, D., Alias, A., Decharme, B., Guérémy, J.-F., Voldoire, A., Abdel-Lathif, A. Y., Bazile, E., Belamari, S., et al. (2020). The cnrm global atmosphere model arpege-climat 6.3: Description and evaluation. *Journal of Advances in Modeling Earth Systems*, 12(7):e2020MS002075.

[Rummukainen, 2016] Rummukainen, M. (2016). Added value in regional climate modeling. *Wiley Interdisciplinary Reviews: Climate Change*, 7(1):145–159.

[Sanchez-Gomez and Somot, 2018] Sanchez-Gomez, E. and Somot, S. (2018). Impact of the internal variability on the cyclone tracks simulated by a regional climate model over the med-cordex domain. *Climate Dynamics*, 51(3):1005–1021.

[Sanchez-Gomez et al., 2009] Sanchez-Gomez, E., Somot, S., and Déqué, M. (2009). Ability of an ensemble of regional climate models to reproduce weather regimes over europe-atlantic during the period 1961–2000. *Climate Dynamics*, 33(5):723–736.

[Schulzweida, 2021] Schulzweida, U. (2021). Cdo user guide.

[Schwingshackl et al., 2019] Schwingshackl, C., Davin, E. L., Hirschi, M., Sørland, S. L., Wartenburger, R., and Seneviratne, S. I. (2019). Regional climate model projections underestimate future warming due to missing plant physiological co2 response. *Environmental Research Letters*, 14(11):114019.

[Séférian et al., 2019] Séférian, R., Nabat, P., Michou, M., Saint-Martin, D., Voldoire, A., Colin, J., Decharme, B., Delire, C., Berthet, S., Chevallier, M., et al. (2019). Evaluation of cnrm earth system model, cnrm-esm2-1: Role of earth system processes in present-day and future climate. *Journal of Advances in Modeling Earth Systems*, 11(12):4182–4227.

[Sieck et al., 2016] Sieck, K., Jacob, D., et al. (2016). Influence of the boundary forcing on the internal variability of a regional climate model. *American Journal of Climate Change*, 5(03):373.

[Sørland et al., 2018] Sørland, S. L., Schär, C., Lüthi, D., and Kjellström, E. (2018). Bias patterns and climate change signals in gcm-rcm model chains. *Environmental Research Letters*, 13(7):074017.

[Staniforth, 1997] Staniforth, A. (1997). Regional modeling: A theoretical discussion. *Meteorology and Atmospheric Physics*, 63(1):15–29.

[Szopa et al., 2013] Szopa, S., Balkanski, Y., Schulz, M., Bekki, S., Cugnet, D., Fortems-Cheiney, A., Turquety, S., Cozic, A., Déandreis, C., Hauglustaine, D., et al. (2013). Aerosol and ozone changes as forcing for climate evolution between 1850 and 2100. *Climate dynamics*, 40(9):2223–2250.

[Taylor et al., 2012] Taylor, K. E., Stouffer, R. J., and Meehl, G. A. (2012). An overview of cmip5 and the experiment design. *Bulletin of the American meteorological Society*, 93(4):485–498.

[Van Vuuren et al., 2011] Van Vuuren, D. P., Edmonds, J., Kainuma, M., Riahi, K., Thomson, A., Hibbard, K., Hurtt, G. C., Kram, T., Krey, V., Lamarque, J.-F., et al. (2011). The representative concentration pathways: an overview. *Climatic change*, 109(1):5–31.

[Vautard et al., 2021] Vautard, R., Kadygrov, N., Iles, C., Boberg, F., Buonomo, E., Bülow, K., Coppola, E., Corre, L., van Meijgaard, E., Nogherotto, R., et al. (2021). Evaluation of the large euro-cordex regional climate model ensemble. *Journal of Geophysical Research: Atmospheres*, 126(17):e2019JD032344.

[Voldoire et al., 2013] Voldoire, A., Sanchez-Gomez, E., Salas y Mélia, D., Decharme, B., Cassou, C., Sénési, S., Valcke, S., Beau, I., Alias, A., Chevallier, M., et al. (2013). The cnrm-cm5. 1 global climate model: description and basic evaluation. *Climate dynamics*, 40(9):2091–2121.

[von Storch et al., 2000] von Storch, H., Langenberg, H., and Feser, F. (2000). A spectral nudging technique for dynamical downscaling purposes. *Monthly weather review*, 128(10):3664–3673.

[Wilks, 2006] Wilks, D. (2006). On "field significance" and the false discovery rate. *Journal of applied meteorology and climatology*, 45(9):1181–1189.

[Wilks, 2016] Wilks, D. (2016). "the stippling shows statistically significant grid points": How research results are routinely overstated and overinterpreted, and what to do about it. *Bulletin of the American Meteorological Society*, 97(12):2263–2273.

Acknowledgements We would like to thank Prof. Dr. Bertrand Denis and Prof. Dr. René Laprise for their feedback on the implementation of 2D-DCT (Discrete Cosine Transform) technique, as well as the development of Big-Brother/Little-Brother experiment which served as inspiration in our work. We would like to thank Dr. Erasmo Buonomo for the fruitful discussion on the assessment of large-scale dynamical consistency of a GCM/RCM pair. We would like to also thank many of our colleagues at the Centre National de Recherches Météorologiques with whom we had numerous exchanges about this research project, and particularly Dr. Bertrand Decharme, Dr. Roland Séférian, Dr. Aurélien Ribes and Dr. Jean-François Guérémy. Finally, we are thankful to the EURO-CORDEX modellers for making the runs available.

Supplementary Material

Wintertime figures

The study deals with the summer season when the strongest GCM-RCM inconsistency are found, as underlined in the introduction. However, for completeness, we also present here the figures for the winter (DJF) season.

Projected winter climate change

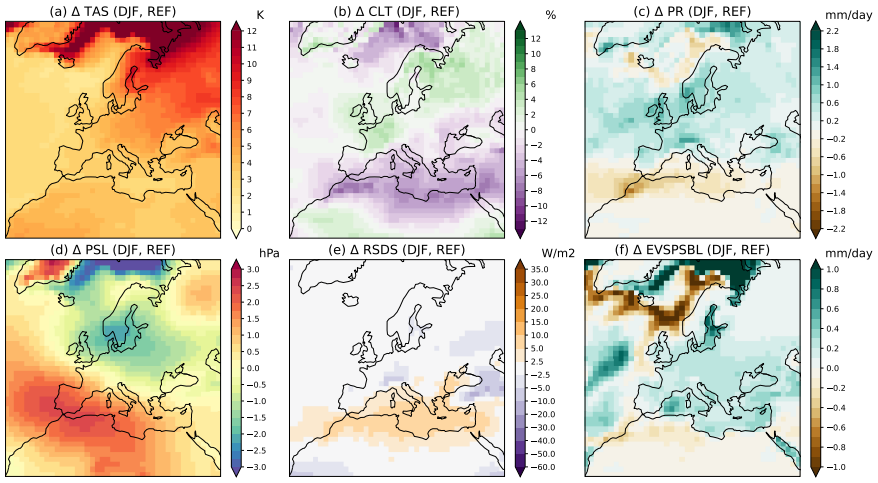

Fig. SM. 1: Mean changes in winter surface temperature (TAS, K), total cloud cover (CLT, %), total precipitation (PR, mm/day), pressure at sea level (PSL, hPa), surface downwelling shortwave radiation (RSDS, W/m2) and evapotranspiration (EVSPSBL, mm/day) as projected by the REF simulation between the RCP8.5 end of century (2070-2099) and historical period (1970-1999).

Historical discrepancies for benchmark experiment REF/ALD150

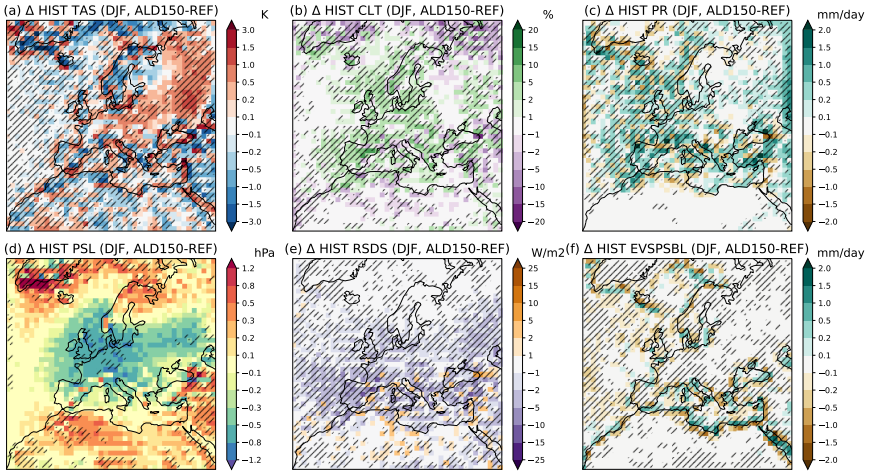

Fig. SM. 2: Mean winter discrepancies between ALD150 and REF for the historical period 1970-1999, computed as the difference between the seasonal means of ALD150-REF. Dashed areas are statistically significant.

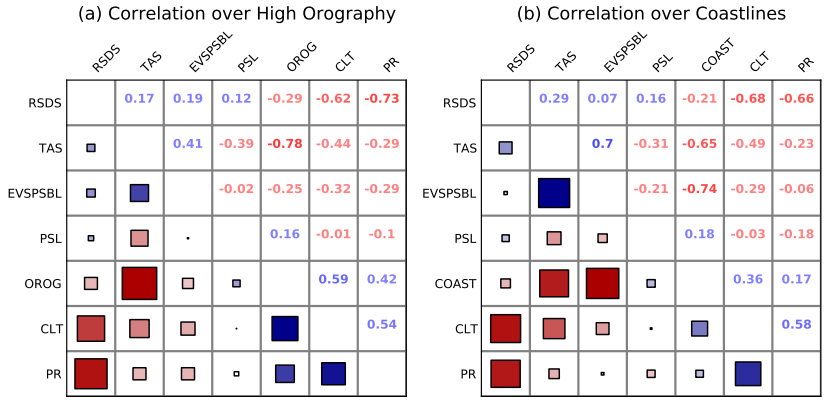

Fig. SM. 3: Pairwise Pearson spatial correlation for winter historical discrepancies between ALD150 and REF. To isolate the potential role of differences in topography (OROG) and coastlines (COAST), spatial correlation was computed over two distinct regions, high topography and coastlines (see details in Article, Sect.2.5). Negative correlations are given in red, and positive correlations in blue. The size of the squares indicate the relative magnitude of the correlations and help navigate faster the chart.

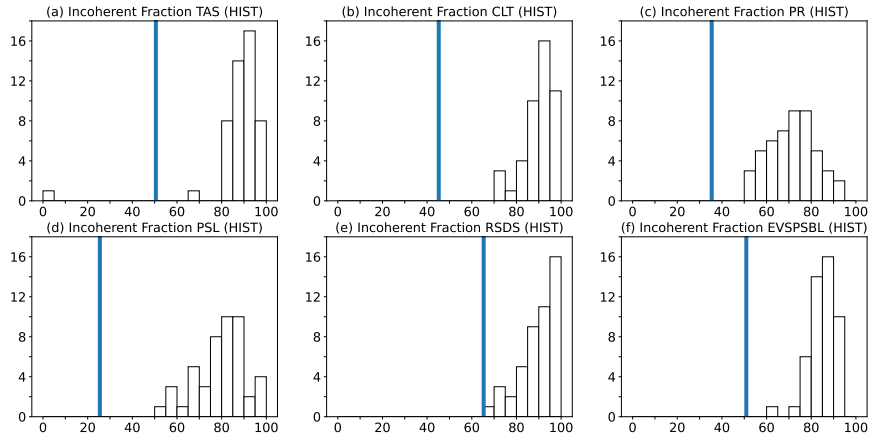

Fig. SM. 4: Fraction of incoherent grid points computed over the land part of the standard EUR-11 domain (black box in Article, Fig.2) for each GCM/RCM pair in the EURO-CORDEX ensemble and REF/ALD150 pair for the historical discrepancies during summer (blue line). The fraction is calculated as the ratio between the number of land grid points where the differences in historical mean climate for a given variable between REF and ALD are statistically significant and the total number of land grid points in the EUR-11 domain. The histograms show the distribution of incoherent fractions in the EURO-CORDEX 48 GCM/RCM pairs.

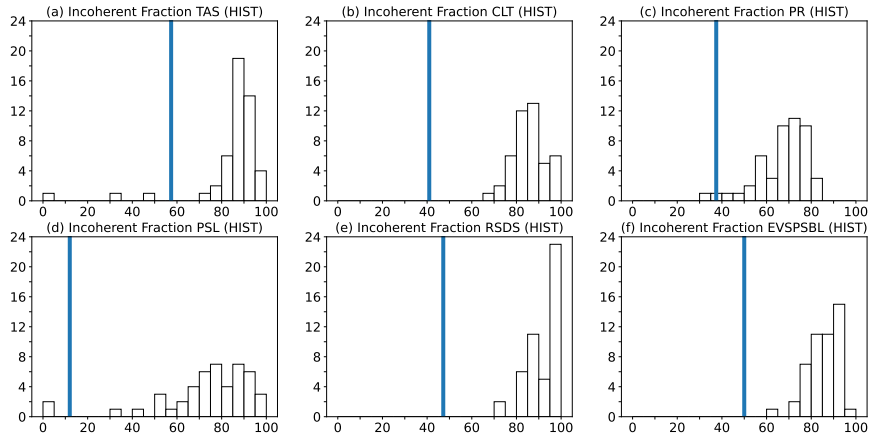

Fig. SM. 5: Fraction of incoherent grid points computed over the land part of the standard EUR-11 domain (black box in Article, Fig.2) for each GCM/RCM pair in the EURO-CORDEX ensemble and REF/ALD150 pair for the historical discrepancies during winter (blue line). The fraction is calculated as the ratio between the number of land grid points where the differences in historical mean climate for a given variable between REF and ALD are statistically significant and the total number of land grid points in the EUR-11 domain. The histograms show the distribution of incoherent fractions in the EURO-CORDEX 48 GCM/RCM pairs.

Differences in projected climate change for all experiments

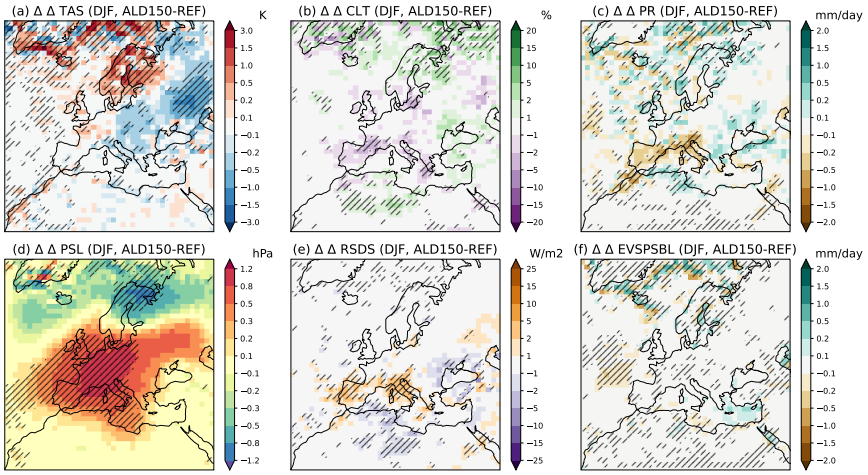

Fig. SM. 6: Mean winter discrepancies in projected climate change signal between ALD150 and REF. Dashed areas are statistically significant (see details in Article Sect. 2.4).

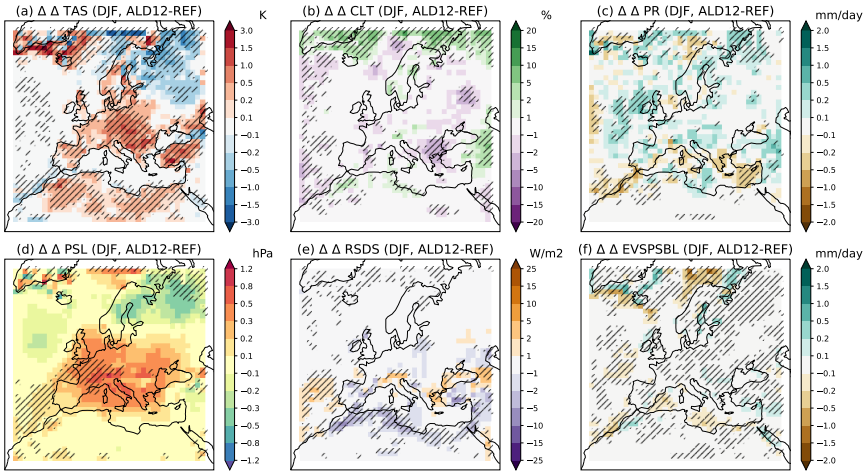

Fig. SM. 7: Mean winter discrepancies in projected climate change signal between ALD12 and REF. Dashed areas are statistically significant (see details in Article Sect. 2.4).

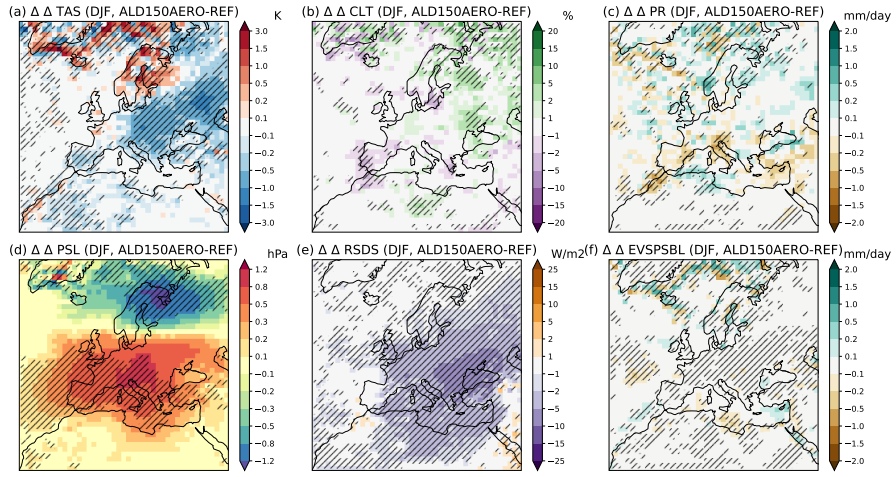

Fig. SM. 8: Mean winter discrepancies in projected climate change signal between ALD150AERO and REF. Dashed areas are statistically significant (see details in Article Sect. 2.4).

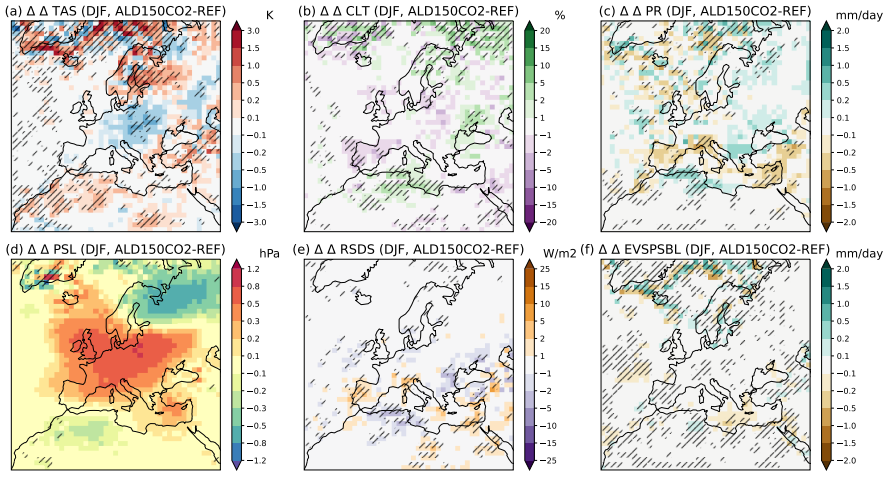

Fig. SM. 9: Mean winter discrepancies in projected climate change signal between ALD150CO2 and REF. Dashed areas are statistically significant (see details in Article Sect. 2.4).

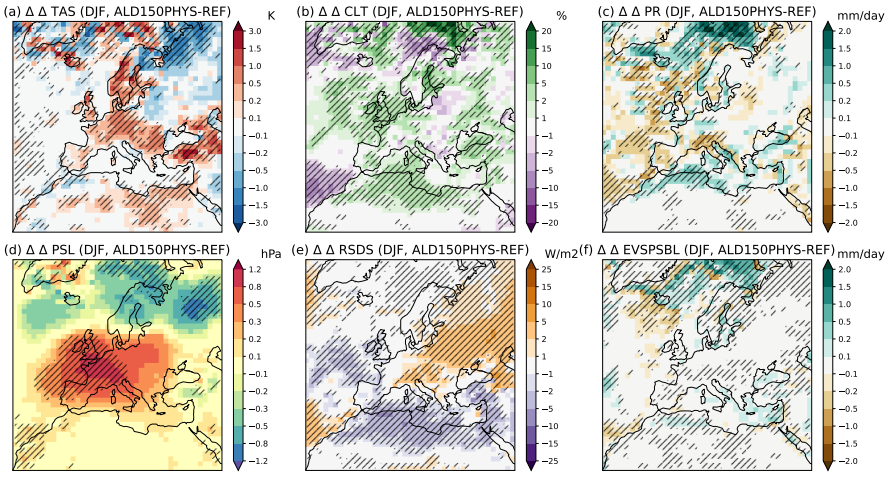

Fig. SM. 10: Mean winter discrepancies in projected climate change signal between ALD150PHYS and REF. Dashed areas are statistically significant (see details in Article Sect. 2.4).

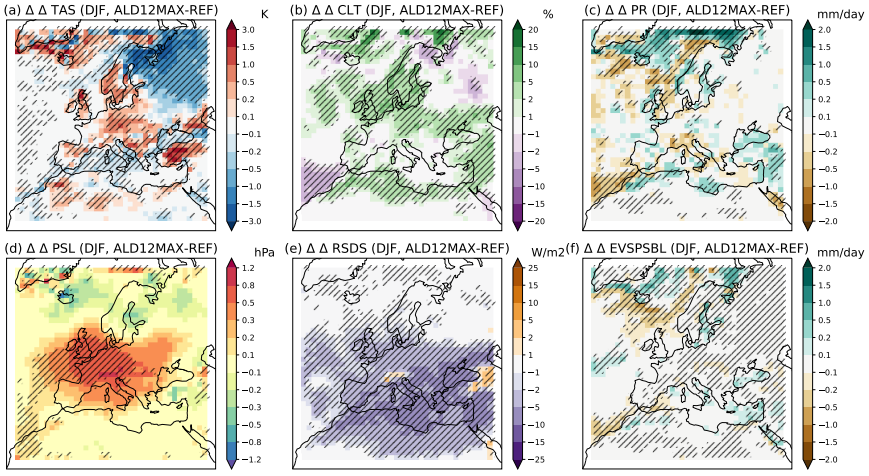

Fig. SM. 11: Mean winter discrepancies in projected climate change signal between ALD12MAX and REF. Dashed areas are statistically significant (see details in Article Sect. 2.4).

Differences in projected climate change (EURO-CORDEX perspective)

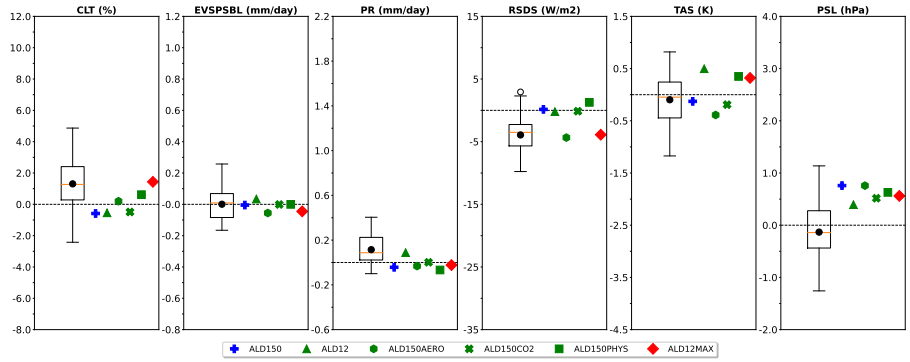

Fig. SM. 12: Mean winter discrepancies in projected climate change over continental Europe, computed over the red box (see Fig. 2) for all EURO-CORDEX GCM/RCM pairs and all experiments presented in the study (detailed methods in Sect. 2.6). The whisker box is representing the range of discrepancies in the EURO-CORDEX ensemble, with black dot corresponding to the mean, orange line to the median, the box covers the interquartile range (IQR) between Q1 and Q3 (25% and 75%), the whiskers covers the interval of Q1-1.5IQR and Q3+1.5IQR, and the outliers are represented as transparent circles. The mean discrepancies corresponding to each experiment are given using a set of colored markers, with a legend provided under the plot.

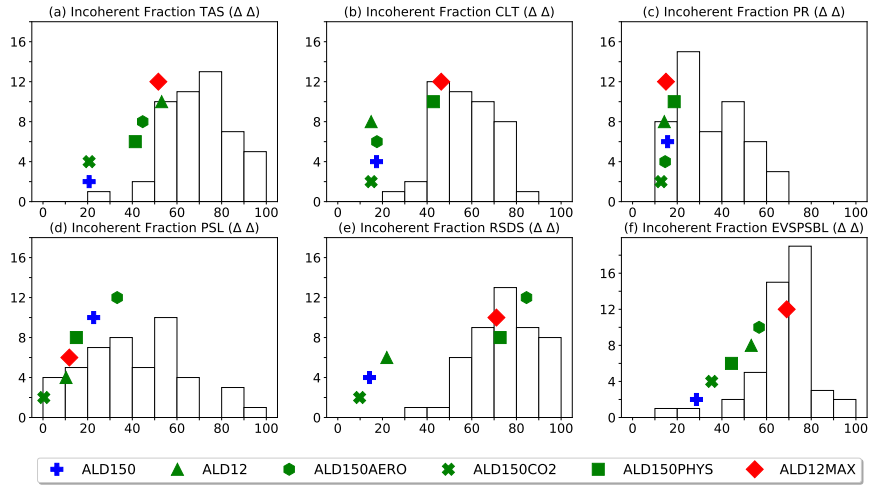

Fig. SM. 13: Fraction of incoherent grid points computed over land part of the standard EUR-11 domain (black box in Article, Fig. 2) for each GCM/RCM pair in the EURO-CORDEX ensemble and each experiment (see methods 2.7) using the discrepancies in winter projected climate change. The fraction is calculated as the ratio between the number of grid points where the differences in projected change for a given variable between REF and ALD are statistically significant and the total number of grid points in the EUR-11 domain. The histograms show the distribution of incoherent fractions in the EURO-CORDEX 48 GCM/RCM pairs, and the fractions for experiments are represented using colored markers with provided legend below the figure.

Impact of each mechanism on the GCM/RCM large scale dynamics

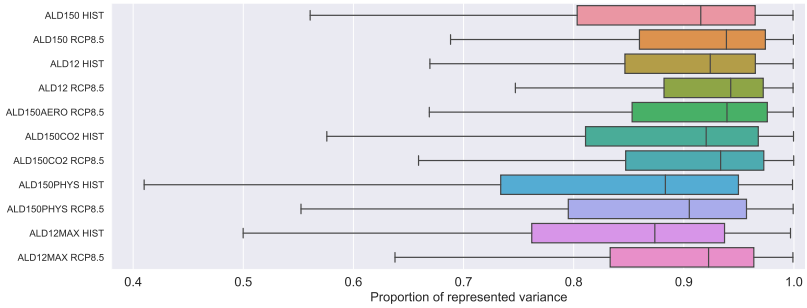

Fig. SM. 14: Box-and-whiskers plot representation of the proportion of represented variance, P(t), computed over the HIST (1970-1999) and RCP8.5 (2070-2099) periods for each experiment for winter. The boxes are delimited by the 25th and 75th percentiles (Q1 and Q3), with the median in between. The range between Q1 and Q3 is the interquartile range (IQR). The minimum and maximum whiskers are given by Q1 - 1.5IQR and Q3+1.5IQR. A value of P(t) close to 1 means that the REF and ALD agree on the large-scale circulation. A P(t) smaller than 1 means that the REF and ALD large-scale circulation differ.

GCM/RCM inconsistencies in the EURO-CORDEX ensemble

Our study uses all the available GCM/RCM pairs for the 12km-resolution EURO-CORDEX scenario simulations for the RCP8.5. To go one step further than the main figures of the article and in order to help the model developers and users, we give below the GCM/RCM inconsistency values for both seasons, every model pair and every variable. In particular, this allows to identify specific GCMs or RCMs or variables for which the inconsistency is generally larger or smaller. We hope that those tables can help to guide future studies to better understand the origin of the GCM/RCM inconsistency.

GCM	Variant	RCM	tas	rsds	pr	evspsbl	clt	psl
CNRM-CERFACS-CNRM-CM5	r1i1p1	CLMcom-CCLM4-8-17	X	X	X	X	X	X
CNRM-CERFACS-CNRM-CM5	r1i1p1	CLMcom-ETH-COSMO-crCLIM-v1-1	X	X	X	X	X	X
CNRM-CERFACS-CNRM-CM5	r1i1p1	CNRM-ALADIN53	X	X	X	X		X
CNRM-CERFACS-CNRM-CM5	r1i1p1	CNRM-ALADIN63	X	X	X	X	X	X
CNRM-CERFACS-CNRM-CM5	r1i1p1	DMI-HIRHAM5	X	X	X	X	X	X
CNRM-CERFACS-CNRM-CM5	r1i1p1	GERICS-REMO2015	X	X	X	X	X	X
CNRM-CERFACS-CNRM-CM5	r1i1p1	ICTP-RegCM4-6	X	X	X	X	X	X
CNRM-CERFACS-CNRM-CM5	r1i1p1	KNMI-RACMO22E	X	X	X	X	X	X
CNRM-CERFACS-CNRM-CM5	r1i1p1	MOHC-HadREM3-GA7-05	X	X	X	X	X	X
CNRM-CERFACS-CNRM-CM5	r1i1p1	RMIB-UGent-ALARO-0	X		X			
CNRM-CERFACS-CNRM-CM5	r1i1p1	SMHI-RCA4	X	X	X	X	X	X
ICHEC-EC-EARTH	r12i1p1	CLMcom-CCLM4-8-17	X	X	X	X	X	X
ICHEC-EC-EARTH	r12i1p1	CLMcom-ETH-COSMO-crCLIM-v1-1	X	X	X	X	X	X
ICHEC-EC-EARTH	r1i1p1	DMI-HIRHAM5	X		X	X		X
ICHEC-EC-EARTH	r12i1p1	ICTP-RegCM4-6	X	X	X	X	X	X
ICHEC-EC-EARTH	r12i1p1	KNMI-RACMO22E	X	X	X	X	X	X
ICHEC-EC-EARTH	r12i1p1	MOHC-HadREM3-GA7-05	X	X	X	X	X	X
ICHEC-EC-EARTH	r12i1p1	SMHI-RCA4	X	X	X	X	X	X
IPSL-IPSL-CM5A-MR	r1i1p1	DMI-HIRHAM5	X	X	X	X	X	X
IPSL-IPSL-CM5A-MR	r1i1p1	GERICS-REMO2015	X	X	X	X	X	X
IPSL-IPSL-CM5A-MR	r1i1p1	IPSL-INERIS-WRF331F	X	X	X	X		
IPSL-IPSL-CM5A-MR	r1i1p1	KNMI-RACMO22E	X	X	X	X	X	X
IPSL-IPSL-CM5A-MR	r1i1p1	SMHI-RCA4	X	X	X	X	X	X
MOHC-HadGEM2-ES	r1i1p1	CLMcom-CCLM4-8-17	X	X	X	X	X	X
MOHC-HadGEM2-ES	r1i1p1	CLMcom-ETH-COSMO-crCLIM-v1-1	X	X	X	X	X	X
MOHC-HadGEM2-ES	r1i1p1	CNRM-ALADIN63	X	X	X	X	X	X
MOHC-HadGEM2-ES	r1i1p1	DMI-HIRHAM5	X	X	X	X	X	X
MOHC-HadGEM2-ES	r1i1p1	ICTP-RegCM4-6	X	X	X	X	X	X
MOHC-HadGEM2-ES	r1i1p1	KNMI-RACMO22E	X	X	X	X	X	X
MOHC-HadGEM2-ES	r1i1p1	MOHC-HadREM3-GA7-05	X	X	X	X	X	X
MOHC-HadGEM2-ES	r1i1p1	SMHI-RCA4	X	X	X	X	X	X
MPI-M-MPI-ESM-LR	r1i1p1	CLMcom-CCLM4-8-17	X	X	X	X	X	X
MPI-M-MPI-ESM-LR	r1i1p1	CLMcom-ETH-COSMO-crCLIM-v1-1	X	X	X	X	X	X
MPI-M-MPI-ESM-LR	r1i1p1	CNRM-ALADIN63	X	X	X	X	X	X
MPI-M-MPI-ESM-LR	r1i1p1	DMI-HIRHAM5	X	X	X	X	X	X
MPI-M-MPI-ESM-LR	r3i1p1	GERICS-REMO2015	X	X	X	X	X	X
MPI-M-MPI-ESM-LR	r1i1p1	ICTP-RegCM4-6	X	X	X	X	X	X
MPI-M-MPI-ESM-LR	r1i1p1	KNMI-RACMO22E	X	X	X	X	X	X
MPI-M-MPI-ESM-LR	r1i1p1	MOHC-HadREM3-GA7-05	X	X	X	X	X	X
MPI-M-MPI-ESM-LR	r1i1p1	MPI-CSC-REMO2009	X	X	X	X	X	X
MPI-M-MPI-ESM-LR	r1i1p1	SMHI-RCA4	X	X	X	X	X	X
NCC-NorESM1-M	r1i1p1	CLMcom-ETH-COSMO-crCLIM-v1-1	X	X	X	X	X	X
NCC-NorESM1-M	r1i1p1	CNRM-ALADIN63	X	X	X	X	X	X
NCC-NorESM1-M	r1i1p1	DMI-HIRHAM5	X	X	X	X	X	X
NCC-NorESM1-M	r1i1p1	GERICS-REMO2015	X	X	X	X	X	X
NCC-NorESM1-M	r1i1p1	ICTP-RegCM4-6	X	X	X	X	X	X
NCC-NorESM1-M	r1i1p1	KNMI-RACMO22E	X	X	X	X	X	X
NCC-NorESM1-M	r1i1p1	MOHC-HadREM3-GA7-05	X	X	X	X	X	X
NCC-NorESM1-M	r1i1p1	SMHI-RCA4	X	X	X	X	X	X

Table. SM. 1: EURO-CORDEX RCMs and their corresponding driving GCMs, with variant label as well as variables used in the study.

GCM	RCM	TAS (JJA)	RSDS (JJA)	PR (JJA)	EVSPSBL (JJA)	CLT (JJA)	PSL (JJA)	Mean (JJA)
CNRM-CERFACS-CNRM-CM5	CLMcom-CCLM4-8-17	68.1	90.2	82.1	87.6	87.1	81.9	82.8
CNRM-CERFACS-CNRM-CM5	CLMcom-ETH-COSMO-crCLIM-v1-1	88.6	87.6	66.1	81.5	99.3	85.4	84.7
CNRM-CERFACS-CNRM-CM5	CNRM-ALADIN63	83.6	77.1	59.6	80.6	90.2	92.3	80.6
CNRM-CERFACS-CNRM-CM5	DMI-HIRHAM5	94.1	91.9	90.8	92.8	90.8	78.6	89.8
CNRM-CERFACS-CNRM-CM5	GERICS-REMO2015	87.8	88.4	79.7	91.5	84.1	80.1	85.3
CNRM-CERFACS-CNRM-CM5	ICTP-RegCM4-6	80.3	99.8	71.8	85.8	91.5	59.2	81.4
CNRM-CERFACS-CNRM-CM5	KNMI-RACMO22E	96.7	86.0	74.2	93.5	91.5	83.8	87.6
CNRM-CERFACS-CNRM-CM5	MOHC-HadREM3-GA7-05	86.5	91.9	79.2	88.2	99.3	76.2	86.9
CNRM-CERFACS-CNRM-CM5	RMIB-UGent-ALARO-0	92.3		78.967				85.6
CNRM-CERFACS-CNRM-CM5	SMHI-RCA4	87.6	83.8	82.3	89.7	85.1	74.2	83.8
ICHEC-EC-EARTH	CLMcom-CCLM4-8-17	84.5	87.3	81.5	90.2	90.0	83.4	86.2
ICHEC-EC-EARTH	CLMcom-ETH-COSMO-crCLIM-v1-1	83.6	80.3	64.6	85.2	94.3	75.3	80.5
ICHEC-EC-EARTH	DMI-HIRHAM5	89.7		76.6	87.3		72.7	81.5
ICHEC-EC-EARTH	ICTP-RegCM4-6	89.9	100.0	57.4	86.5	92.8	69.4	82.7
ICHEC-EC-EARTH	KNMI-RACMO22E	95.2	85.6	53.5	81.9	81.5	85.8	80.6
ICHEC-EC-EARTH	MOHC-HadREM3-GA7-05	81.9	72.9	68.8	77.3	88.4	87.3	79.4
ICHEC-EC-EARTH	SMHI-RCA4	88.0	92.6	66.2	82.5	90.8	97.4	86.3
IPSL-IPSL-CM5A-MR	DMI-HIRHAM5	96.9	99.6	56.3	77.3	92.4	79.7	83.7
IPSL-IPSL-CM5A-MR	GERICS-REMO2015	95.0	99.4	67.5	79.5	98.2	67.3	84.5
IPSL-IPSL-CM5A-MR	IPSL-INERIS-WRF331F	97.0	86.7	64.6	83.9			83.1
IPSL-IPSL-CM5A-MR	KNMI-RACMO22E	95.2	99.1	63.5	90.4	98.3	79.3	87.6
IPSL-IPSL-CM5A-MR	SMHI-RCA4	94.3	96.3	77.5	90.2	95.8	91.9	91.0
MOHC-HadGEM2-ES	CLMcom-CCLM4-8-17	87.6	99.8	92.3	87.3	97.0	88.9	92.2
MOHC-HadGEM2-ES	CLMcom-ETH-COSMO-crCLIM-v1-1	93.7	92.6	71.2	84.3	97.8	96.1	89.3
MOHC-HadGEM2-ES	CNRM-ALADIN63	91.0	89.5	51.1	83.4	94.8	84.5	82.4
MOHC-HadGEM2-ES	DMI-HIRHAM5	95.2	99.8	75.6	88.7	91.3	81.2	88.7
MOHC-HadGEM2-ES	ICTP-RegCM4-6	82.7	96.1	51.1	84.9	71.8	76.0	77.1
MOHC-HadGEM2-ES	KNMI-RACMO22E	92.4	94.8	63.8	85.1	93.9	86.2	86.0
MOHC-HadGEM2-ES	MOHC-HadREM3-GA7-05	91.3	88.4	68.5	80.3	98.7	85.6	85.5
MOHC-HadGEM2-ES	SMHI-RCA4	94.1	86.2	74.0	88.9	85.4	95.9	87.4
MPI-M-MPI-ESM-LR	CLMcom-CCLM4-8-17	91.1	90.2	58.1	80.6	93.0	95.2	84.7
MPI-M-MPI-ESM-LR	CLMcom-ETH-COSMO-crCLIM-v1-1	87.8	76.4	65.9	78.0	89.9	67.5	77.6
MPI-M-MPI-ESM-LR	CNRM-ALADIN63	88.0	96.5	80.1	88.6	81.4	87.3	87.0
MPI-M-MPI-ESM-LR	DMI-HIRHAM5	91.7	90.0	64.0	62.2	78.8	63.3	75.0
MPI-M-MPI-ESM-LR	GERICS-REMO2015	86.5	81.2	74.9	76.2	89.5	68.8	79.5
MPI-M-MPI-ESM-LR	ICTP-RegCM4-6	84.1	99.8	77.1	90.8	89.3	74.0	85.9
MPI-M-MPI-ESM-LR	KNMI-RACMO22E	93.4	74.5	75.1	94.5	72.7	55.2	77.6
MPI-M-MPI-ESM-LR	MOHC-HadREM3-GA7-05	82.3	91.0	85.2	88.7	84.1	85.2	86.1
MPI-M-MPI-ESM-LR	MPI-CSC-REMO2009	86.9	80.4	77.1	81.2	86.3	69.0	80.2
MPI-M-MPI-ESM-LR	SMHI-RCA4	92.6	95.9	84.1	92.4	90.4	82.8	89.7
NCC-NorESM1-M	CLMcom-ETH-COSMO-crCLIM-v1-1	94.6	93.2	72.0	80.6	99.3	79.9	86.6
NCC-NorESM1-M	CNRM-ALADIN63	88.7	83.2	72.5	88.6	96.9	89.1	86.5
NCC-NorESM1-M	DMI-HIRHAM5	95.2	98.2	87.6	90.0	92.8	86.7	91.8
NCC-NorESM1-M	GERICS-REMO2015	90.6	95.0	86.0	85.4	85.4	77.1	86.6
NCC-NorESM1-M	ICTP-RegCM4-6	85.4	99.6	60.9	86.2	71.8	57.4	76.9
NCC-NorESM1-M	KNMI-RACMO22E	93.9	96.1	58.3	83.4	92.4	83.2	84.6
NCC-NorESM1-M	MOHC-HadREM3-GA7-05	92.3	90.8	70.5	79.3	98.3	83.9	85.9
NCC-NorESM1-M	SMHI-RCA4	92.3	74.7	72.0	82.3	87.8	82.8	82.0

	Value
	[0,20)
	[20,40)
	[40,60)
	[60,80)
	[80,100]

Table. SM. 2: Fraction of incoherent grid points computed over land part of the standard EUR-11 domain (black box in Article, Fig. 2) for each GCM/RCM pair in the EURO-CORDEX ensemble (see methods 2.7) using the summer discrepancies of the historical period (1970-1999). The fraction is calculated as the ratio between the number of land points where the differences in historical mean for a given variable between REF and ALD are statistically significant and the total number of land points in the EUR-11 domain.

GCM	RCM	TAS (DJF)	RSDS (DFJ)	PR (DJF)	EVSPSBL (DJF)	CLT (DJF)	PSL (DJF)	MEAN (DJF)
CNRM-CERFACS-CNRM-CM5	CLMcom-CCLM4-8-17	46.9	84.9	46.9	83.4	78.0	43.0	63.8
CNRM-CERFACS-CNRM-CM5	CLMcom-ETH-COSMO-crCLIM-v1-1	85.2	93.2	61.6	83.2	91.5	100.0	85.8
CNRM-CERFACS-CNRM-CM5	CNRM-ALADIN63	81.4	97.8	69.9	84.1	95.4	72.1	83.5
CNRM-CERFACS-CNRM-CM5	DMI-HIRHAM5	85.2	99.3	78.8	86.5	85.6	51.8	81.2
CNRM-CERFACS-CNRM-CM5	GERICS-REMO2015	81.4	99.3	56.8	76.2	77.9	67.0	76.4
CNRM-CERFACS-CNRM-CM5	ICTP-RegCM4-6	77.7	96.7	66.6	88.4	85.1	76.8	81.9
CNRM-CERFACS-CNRM-CM5	KNMI-RACMO22E	87.3	86.5	53.0	83.0	91.5	72.7	79.0
CNRM-CERFACS-CNRM-CM5	MOHC-HadREM3-GA7-05	82.7	88.2	72.3	81.4	98.5	72.7	82.6
CNRM-CERFACS-CNRM-CM5	RMIB-UGent-ALARO-0	70.1		36.2				53.1
CNRM-CERFACS-CNRM-CM5	SMHI-RCA4	31.5	99.8	34.5	64.6	75.5	2.8	51.4
ICHEC-EC-EARTH	CLMcom-CCLM4-8-17	94.3	93.9	63.1	85.1	81.9	99.8	86.3
ICHEC-EC-EARTH	CLMcom-ETH-COSMO-crCLIM-v1-1	86.5	99.4	71.6	73.8	87.6	99.6	86.4
ICHEC-EC-EARTH	DMI-HIRHAM5	90.0		77.9	97.0		85.6	87.6
ICHEC-EC-EARTH	ICTP-RegCM4-6	90.4	100.0	76.6	83.8	84.3	68.6	83.9
ICHEC-EC-EARTH	KNMI-RACMO22E	99.6	96.7	53.3	86.9	77.7	89.9	84.0
ICHEC-EC-EARTH	MOHC-HadREM3-GA7-05	92.6	81.7	73.8	75.6	96.9	67.7	81.4
ICHEC-EC-EARTH	SMHI-RCA4	95.2	99.3	65.3	76.4	91.0	92.8	86.7
IPSL-IPSL-CM5A-MR	DMI-HIRHAM5	88.4	95.0	57.6	84.9	82.3	89.5	82.9
IPSL-IPSL-CM5A-MR	GERICS-REMO2015	86.7	96.7	66.4	85.8	89.1	74.2	83.1
IPSL-IPSL-CM5A-MR	IPSL-INERIS-WRF331F	94.5	97.4	67.3	87.8			86.8
IPSL-IPSL-CM5A-MR	KNMI-RACMO22E	95.4	89.5	72.0	88.9	84.1	85.1	85.8
IPSL-IPSL-CM5A-MR	SMHI-RCA4	77.5	95.0	68.8	87.8	89.5	90.8	84.9
MOHC-HadGEM2-ES	CLMcom-CCLM4-8-17	86.3	86.9	73.6	91.0	86.0	79.3	83.9
MOHC-HadGEM2-ES	CLMcom-ETH-COSMO-crCLIM-v1-1	93.2	81.4	78.0	87.5	84.9	68.5	82.2
MOHC-HadGEM2-ES	CNRM-ALADIN63	90.4	86.2	56.8	80.1	89.1	54.8	76.2
MOHC-HadGEM2-ES	DMI-HIRHAM5	97.0	100.0	80.4	94.1	86.7	81.9	90.0
MOHC-HadGEM2-ES	ICTP-RegCM4-6	90.2	89.7	67.5	76.6	90.6	87.8	83.7
MOHC-HadGEM2-ES	KNMI-RACMO22E	83.8	88.7	55.7	91.3	84.5	70.3	79.1
MOHC-HadGEM2-ES	MOHC-HadREM3-GA7-05	86.9	83.4	73.4	82.1	98.0	83.2	84.5
MOHC-HadGEM2-ES	SMHI-RCA4	92.8	89.3	82.7	78.0	81.9	77.3	83.7
MPI-M-MPI-ESM-LR	CLMcom-CCLM4-8-17	92.3	81.4	71.4	91.5	70.3	81.9	81.5
MPI-M-MPI-ESM-LR	CLMcom-ETH-COSMO-crCLIM-v1-1	87.5	80.6	65.9	92.4	79.2	91.3	82.8
MPI-M-MPI-ESM-LR	CNRM-ALADIN63	86.7	89.7	73.8	93.2	84.9	72.9	83.5
MPI-M-MPI-ESM-LR	DMI-HIRHAM5	85.1	98.5	75.8	92.6	81.0	57.6	81.8
MPI-M-MPI-ESM-LR	GERICS-REMO2015	93.4	98.5	74.4	90.6	80.4	55.0	82.0
MPI-M-MPI-ESM-LR	ICTP-RegCM4-6	85.6	93.4	83.6	91.7	79.5	33.6	77.9
MPI-M-MPI-ESM-LR	KNMI-RACMO22E	94.1	85.6	76.9	94.8	85.6	75.8	85.5
MPI-M-MPI-ESM-LR	MOHC-HadREM3-GA7-05	86.3	74.0	76.2	91.3	96.5	86.9	85.2
MPI-M-MPI-ESM-LR	MPI-CSC-REMO2009	92.8	98.7	75.1	92.3	73.8	63.8	82.7
MPI-M-MPI-ESM-LR	SMHI-RCA4	85.6	96.3	70.7	91.3	85.2	62.5	81.9
NCC-NorESM1-M	CLMcom-ETH-COSMO-crCLIM-v1-1	85.6	97.4	69.0	83.9	88.6	94.1	86.4
NCC-NorESM1-M	CNRM-ALADIN63	89.9	96.5	60.5	70.5	91.3	90.4	83.2
NCC-NorESM1-M	DMI-HIRHAM5	89.3	96.3	69.2	92.3	87.8	81.9	86.1
NCC-NorESM1-M	GERICS-REMO2015	85.1	93.9	59.8	91.5	81.2	79.9	81.9
NCC-NorESM1-M	ICTP-RegCM4-6	83.6	99.4	76.9	88.9	67.5	77.9	82.4
NCC-NorESM1-M	KNMI-RACMO22E	93.7	86.0	76.6	86.7	88.2	85.1	86.0
NCC-NorESM1-M	MOHC-HadREM3-GA7-05	81.2	92.8	60.0	79.7	97.8	90.8	83.7
NCC-NorESM1-M	SMHI-RCA4	85.2	98.9	71.4	81.2	83.0	76.4	82.7

	Value
	[0,20)
	[20,40)
	[40,60)
	[60,80)
	[80,100]

Table. SM. 3: Same as Table SM. 2, but for winter.

GCM	RCM	TAS (JJA)	RSDS (JJA)	PR (JJA)	EVSPSBL (JJA)	CLT (JJA)	PSL (JJA)	Mean (JJA)
CNRM-CERFACS-CNRM-CM5	CLMcom-CCLM4-8-17	95.2	94.8	15.9	48.2	51.1	81.5	64.5
CNRM-CERFACS-CNRM-CM5	CLMcom-ETH-COSMO-crCLIM-v1-1	90.4	95.8	21.8	60.7	50.0	39.7	59.7
CNRM-CERFACS-CNRM-CM5	CNRM-ALADIN63	30.4	29.5	12.4	42.4	36.5	51.7	33.8
CNRM-CERFACS-CNRM-CM5	DMI-HIRHAM5	74.7	84.9	13.7	34.9	49.1	44.1	50.2
CNRM-CERFACS-CNRM-CM5	GERICS-REMO2015	67.0	90.4	14.8	54.1	19.2	6.5	42.0
CNRM-CERFACS-CNRM-CM5	ICTP-RegCM4-6	32.1	76.6	18.1	72.0	63.7	30.8	48.9
CNRM-CERFACS-CNRM-CM5	KNMI-RACMO22E	56.5	78.2	8.5	32.7	26.6	18.8	36.9
CNRM-CERFACS-CNRM-CM5	MOHC-HadREM3-GA7-05	81.4	71.6	40.0	64.0	93.5	97.8	74.7
CNRM-CERFACS-CNRM-CM5	RMIB-UGent-ALARO-0	55.0		16.8				35.9
CNRM-CERFACS-CNRM-CM5	SMHI-RCA4	82.5	95.2	12.7	51.1	39.5	22.1	50.5
ICHEC-EC-EARTH	CLMcom-CCLM4-8-17	79.9	87.8	28.2	50.0	56.8	42.3	57.5
ICHEC-EC-EARTH	CLMcom-ETH-COSMO-crCLIM-v1-1	81.5	88.7	13.5	31.9	55.5	22.3	48.9
ICHEC-EC-EARTH	DMI-HIRHAM5	71.6		16.8	36.2		25.8	37.6
ICHEC-EC-EARTH	ICTP-RegCM4-6	34.9	77.1	20.7	47.8	29.7	54.8	44.2
ICHEC-EC-EARTH	KNMI-RACMO22E	71.2	40.0	25.5	21.6	33.0	41.1	38.7
ICHEC-EC-EARTH	MOHC-HadREM3-GA7-05	86.0	95.6	32.5	49.8	94.6	99.6	76.4
ICHEC-EC-EARTH	SMHI-RCA4	37.1	85.8	23.1	55.4	51.1	17.5	45.0
IPSL-IPSL-CM5A-MR	DMI-HIRHAM5	81.7	78.2	62.9	86.9	72.0	46.3	71.3
IPSL-IPSL-CM5A-MR	GERICS-REMO2015	84.9	91.5	61.1	61.6	74.5	46.3	70.0
IPSL-IPSL-CM5A-MR	IPSL-INERIS-WRF331F	63.8	50.2	44.3	65.5			56.0
IPSL-IPSL-CM5A-MR	KNMI-RACMO22E	69.0	74.2	51.1	84.5	59.0	37.3	62.5
IPSL-IPSL-CM5A-MR	SMHI-RCA4	71.4	86.7	57.9	78.8	78.2	74.0	74.5
MOHC-HadGEM2-ES	CLMcom-CCLM4-8-17	94.8	87.3	64.8	77.1	83.4	100.0	84.6
MOHC-HadGEM2-ES	CLMcom-ETH-COSMO-crCLIM-v1-1	91.1	93.5	61.4	83.2	95.8	86.5	85.3
MOHC-HadGEM2-ES	CNRM-ALADIN63	78.6	91.7	58.5	74.4	89.5	82.7	79.2
MOHC-HadGEM2-ES	DMI-HIRHAM5	98.9	89.5	59.0	81.2	89.5	86.5	84.1
MOHC-HadGEM2-ES	ICTP-RegCM4-6	88.7	88.2	74.2	79.5	87.5	85.2	83.9
MOHC-HadGEM2-ES	KNMI-RACMO22E	79.3	73.1	64.0	91.9	81.2	90.2	80.0
MOHC-HadGEM2-ES	MOHC-HadREM3-GA7-05	62.2	79.9	44.6	67.0	77.9	69.0	66.8
MOHC-HadGEM2-ES	SMHI-RCA4	83.6	87.5	67.9	77.3	87.8	83.8	81.3
MPI-M-MPI-ESM-LR	CLMcom-CCLM4-8-17	78.0	84.1	50.4	78.2	56.3	64.6	68.6
MPI-M-MPI-ESM-LR	CLMcom-ETH-COSMO-crCLIM-v1-1	71.4	71.6	41.5	68.6	58.9	47.0	59.8
MPI-M-MPI-ESM-LR	CNRM-ALADIN63	51.1	52.4	31.9	65.9	43.0	64.9	51.5
MPI-M-MPI-ESM-LR	DMI-HIRHAM5	67.7	60.5	47.6	80.1	42.3	37.6	56.0
MPI-M-MPI-ESM-LR	GERICS-REMO2015	75.5	82.3	46.7	53.7	69.2	46.3	62.3
MPI-M-MPI-ESM-LR	ICTP-RegCM4-6	74.9	51.8	43.5	53.5	51.5	62.9	56.4
MPI-M-MPI-ESM-LR	KNMI-RACMO22E	79.0	55.7	48.3	91.9	63.7	42.1	63.4
MPI-M-MPI-ESM-LR	MOHC-HadREM3-GA7-05	77.3	91.5	38.4	64.8	68.1	95.4	72.6
MPI-M-MPI-ESM-LR	MPI-CSC-REMO2009	75.3	75.3	47.6	64.0	64.6	35.2	60.3
MPI-M-MPI-ESM-LR	SMHI-RCA4	66.8	64.8	40.0	65.3	69.9	40.6	57.9
NCC-NorESM1-M	CLMcom-ETH-COSMO-crCLIM-v1-1	85.8	77.7	25.1	54.8	53.3	71.8	61.4
NCC-NorESM1-M	CNRM-ALADIN63	44.5	32.5	28.0	44.6	35.4	37.1	37.0
NCC-NorESM1-M	DMI-HIRHAM5	98.7	75.6	32.7	67.2	59.8	97.2	71.9
NCC-NorESM1-M	GERICS-REMO2015	92.8	93.9	32.3	38.6	22.5	56.8	56.2
NCC-NorESM1-M	ICTP-RegCM4-6	75.6	91.7	13.8	49.3	42.8	80.3	58.9
NCC-NorESM1-M	KNMI-RACMO22E	72.9	60.1	33.0	67.2	47.4	83.8	60.7
NCC-NorESM1-M	MOHC-HadREM3-GA7-05	59.6	83.4	38.7	58.3	93.5	42.8	62.7
NCC-NorESM1-M	SMHI-RCA4	77.3	83.9	32.3	54.2	36.7	82.8	61.2

Value
[0,20)
[20,40)
[40,60)
[60,80)
[80,100]

Table. SM. 4: Fraction of incoherent grid points computed over land part of the standard EUR-11 domain (black box in Article, Fig. 2) for each GCM/RCM pair in the EURO-CORDEX ensemble (see methods 2.7) using the summer discrepancies in projected climate change between RCP8.5 (2070-2099) and historical (1970-1999) periods. The fraction is calculated as the ratio between the number of land points where the differences in projected change for a given variable between REF and ALD are statistically significant and the total number of land points in the EUR-11 domain.

GCM	RCM	TAS (DJF)	RSDS (DJF)	PR (DJF)	EVSPSBL (DJF)	CLT (DJF)	PSL (DJF)	Mean (DJF)
CNRM-CERFACS-CNRM-CM5	CLMcom-CCLM4-8-17	91.3	95.0	31.4	74.9	58.3	50.7	66.9
CNRM-CERFACS-CNRM-CM5	CLMcom-ETH-COSMO-crCLIM-v1-1	88.4	98.9	17.9	77.3	47.8	22.0	58.7
CNRM-CERFACS-CNRM-CM5	CNRM-ALADIN63	54.2	74.5	40.8	69.4	49.6	37.1	54.3
CNRM-CERFACS-CNRM-CM5	DMI-HIRHAM5	47.6	67.9	43.2	73.1	60.5	19.0	51.9
CNRM-CERFACS-CNRM-CM5	GERICS-REMO2015	52.0	63.5	30.1	65.5	34.5	31.4	46.2
CNRM-CERFACS-CNRM-CM5	ICTP-RegCM4-6	29.9	75.6	21.0	70.1	54.8	10.3	43.6
CNRM-CERFACS-CNRM-CM5	KNMI-RACMO22E	61.6	74.5	21.4	68.3	50.4	50.9	54.5
CNRM-CERFACS-CNRM-CM5	MOHC-HadREM3-GA7-05	57.9	90.0	43.7	61.8	63.7	29.5	57.8
CNRM-CERFACS-CNRM-CM5	RMIB-UGent-ALARO-0	84.7		16.8				50.7
CNRM-CERFACS-CNRM-CM5	SMHI-RCA4	59.8	86.9	46.1	64.6	47.8	53.0	59.7
ICHEC-EC-EARTH	CLMcom-CCLM4-8-17	58.7	98.2	18.1	50.6	36.7	41.7	50.6
ICHEC-EC-EARTH	CLMcom-ETH-COSMO-crCLIM-v1-1	58.9	92.4	22.1	62.4	28.6	41.1	50.9
ICHEC-EC-EARTH	DMI-HIRHAM5	55.4		29.3	67.7		6.8	39.8
ICHEC-EC-EARTH	ICTP-RegCM4-6	62.5	48.5	11.8	57.7	44.3	21.0	41.0
ICHEC-EC-EARTH	KNMI-RACMO22E	73.8	32.5	14.6	48.5	44.3	44.6	43.1
ICHEC-EC-EARTH	MOHC-HadREM3-GA7-05	85.8	82.1	28.4	19.0	59.4	60.7	55.9
ICHEC-EC-EARTH	SMHI-RCA4	77.1	84.5	18.1	23.2	48.2	9.2	43.4
IPSL-IPSL-CM5A-MR	DMI-HIRHAM5	64.4	71.4	60.9	77.1	79.7	39.1	65.4
IPSL-IPSL-CM5A-MR	GERICS-REMO2015	67.5	81.2	43.4	71.6	76.6	51.8	65.3
IPSL-IPSL-CM5A-MR	IPSL-INERIS-WRF331F	74.0	68.6	46.7	68.8			64.5
IPSL-IPSL-CM5A-MR	KNMI-RACMO22E	81.7	77.1	45.4	79.3	77.1	38.7	66.6
IPSL-IPSL-CM5A-MR	SMHI-RCA4	81.2	62.2	59.6	77.9	75.5	56.3	68.8
MOHC-HadGEM2-ES	CLMcom-CCLM4-8-17	89.9	98.9	39.5	73.2	60.3	92.8	75.8
MOHC-HadGEM2-ES	CLMcom-ETH-COSMO-crCLIM-v1-1	95.8	95.6	32.7	71.2	60.1	53.0	68.1
MOHC-HadGEM2-ES	CNRM-ALADIN63	79.3	55.7	18.3	53.5	56.8	39.7	50.6
MOHC-HadGEM2-ES	DMI-HIRHAM5	97.6	57.4	28.8	71.6	64.6	27.1	57.8
MOHC-HadGEM2-ES	ICTP-RegCM4-6	76.6	59.0	21.4	59.0	42.3	8.1	44.4
MOHC-HadGEM2-ES	KNMI-RACMO22E	63.5	54.2	18.3	64.8	67.5	27.1	49.2
MOHC-HadGEM2-ES	MOHC-HadREM3-GA7-05	77.3	60.9	34.5	60.7	76.2	16.2	54.3
MOHC-HadGEM2-ES	SMHI-RCA4	75.3	85.1	23.2	47.6	78.4	38.0	57.9
MPI-M-MPI-ESM-LR	CLMcom-CCLM4-8-17	70.8	80.1	50.9	84.9	43.7	60.5	65.2
MPI-M-MPI-ESM-LR	CLMcom-ETH-COSMO-crCLIM-v1-1	84.1	82.1	55.4	78.2	67.7	59.0	71.1
MPI-M-MPI-ESM-LR	CNRM-ALADIN63	75.6	81.0	56.1	91.0	64.0	57.2	70.8
MPI-M-MPI-ESM-LR	DMI-HIRHAM5	74.0	62.5	43.2	73.1	43.4	30.8	54.5
MPI-M-MPI-ESM-LR	GERICS-REMO2015	62.5	58.5	30.4	65.5	63.7	20.8	50.2
MPI-M-MPI-ESM-LR	ICTP-RegCM4-6	68.8	72.9	63.5	77.5	43.5	17.2	57.2
MPI-M-MPI-ESM-LR	KNMI-RACMO22E	76.9	68.6	57.0	81.9	50.4	83.9	69.8
MPI-M-MPI-ESM-LR	MOHC-HadREM3-GA7-05	70.8	71.6	46.3	86.0	70.7	81.2	71.1
MPI-M-MPI-ESM-LR	MPI-CSC-REMO2009	66.6	72.5	56.1	76.8	66.1	62.0	66.7
MPI-M-MPI-ESM-LR	SMHI-RCA4	69.4	75.5	67.0	91.3	82.7	49.6	72.6
NCC-NorESM1-M	CLMcom-ETH-COSMO-crCLIM-v1-1	95.4	99.3	25.6	74.0	46.5	61.3	67.0
NCC-NorESM1-M	CNRM-ALADIN63	61.3	53.7	28.8	66.1	56.1	53.7	53.3
NCC-NorESM1-M	DMI-HIRHAM5	95.6	74.2	43.0	72.9	58.7	43.4	64.6
NCC-NorESM1-M	GERICS-REMO2015	40.4	77.5	24.2	72.9	40.0	7.2	43.7
NCC-NorESM1-M	ICTP-RegCM4-6	56.8	78.8	29.2	67.9	50.6	17.5	50.1
NCC-NorESM1-M	KNMI-RACMO22E	63.3	64.2	25.3	77.3	57.2	53.7	56.8
NCC-NorESM1-M	MOHC-HadREM3-GA7-05	52.6	61.1	23.4	51.7	75.3	39.1	50.5
NCC-NorESM1-M	SMHI-RCA4	50.2	87.8	26.4	64.0	55.7	23.6	51.3

Value
[0,20)
[20,40)
[40,60)
[60,80)
[80,100]

Table. SM. 5: Same as Table SM. 4, but for winter.

	ICHEC-EC-EARTH	MOHC-HadGEM2-ES	MPI-M-MPI-ESM-LR	NCC-NorESM1-M	CNRM-CERFACS-CNRM-CM5	IPSL-IPSL-CM5A-MR	MEAN (RCM)
CLMcom-CCLM4-8-17	57.5	84.0	68.6	61.4	64.5	?	67.3
CLMcom-ETH-COSMO-crCLIM-v1-1	48.9	85.3	59.8	?	59.7	?	63.4
CNRM-ALADIN63	?	79.2	51.5	37.0	33.8	?	50.4
DMI-HIRHAM5	?	84.1	56.0	71.9	50.2	71.3	66.7
GERICS-REMO2015	?	?	62.3	56.2	42.0	70.0	57.6
ICTP-RegCM4-6	44.2	83.9	56.4	58.9	48.9	?	58.4
KNMI-RACMO22E	38.7	80.0	63.4	60.7	36.9	62.5	57.0
MOHC-HadREM3-GA7-05	76.4	66.8	72.6	62.7	74.7	?	70.6
SMHI-RCA4	45.0	81.3	57.9	61.2	50.5	74.5	61.7
MEAN (GCM)	51.8	80.6	60.9	58.8	51.2	69.6	

Theoretical Best	
REF/ALD150	16.4
REF/ALD12	27.4

Value
[0,20)
[20,40)
[40,60)
[60,80)
[80,100]

Table. SM. 6: Mean fraction of incoherent grid points computed over land part of the standard EUR-11 domain (black box in Article, Fig. 2) for each GCM/RCM pair in the EURO-CORDEX ensemble (see methods 2.7) using the summer discrepancies in projected climate change between RCP8.5 (2070-2099) and historical (1970-1999) periods for all six studied variables. The mean fraction is calculated as the average over all variables from Table SM. 4.

	ICHEC-EC-EARTH	MOHC-HadGEM2-ES	MPI-M-MPI-ESM-LR	NCC-NorESM1-M	CNRM-CERFACS-CNRM-CM5	IPSL-IPSL-CM5A-MR	MEAN (RCM)
CLMcom-CCLM4-8-17	50.6	75.8	65.2	67.0	66.9	?	65.1
CLMcom-ETH-COSMO-crCLIM-v1-1	50.9	68.1	71.1	?	58.7	?	62.2
CNRM-ALADIN63	?	50.6	70.8	53.3	54.3	?	57.2
DMI-HIRHAM5	?	57.8	54.5	64.6	51.9	65.4	58.8
GERICS-REMO2015	?	?	50.2	43.7	46.2	65.3	51.4
ICTP-RegCM4-6	41.0	44.4	57.2	50.1	43.6	?	47.3
KNMI-RACMO22E	43.1	49.2	69.8	56.8	54.5	66.6	56.7
MOHC-HadREM3-GA7-05	55.9	54.3	71.1	50.5	57.8	?	57.9
SMHI-RCA4	43.4	57.9	72.6	51.3	59.7	68.8	58.9
MEAN (GCM)	47.5	57.3	60.9	54.7	54.8	66.5	

Theoretical Best	
REF/ALD150	19.9
REF/ALD12	28.0

	Value
	[0,20)
	[20,40)
	[40,60)
	[60,80)
	[80,100]

Table. SM. 7: Same as Table SM. 6, but for winter.

GCM	RCM	TAS (K)	RSDS (W/m2)	PR (mm/day)	EVSPSBL (mm/day)	CLT (%)	PSL (hPa)
CNRM-CERFACS-CNRM-CM5	CLMcom-CCLM4-8-17	-1.3	-28.5	-0.1	-0.2	-1.3	0.6
CNRM-CERFACS-CNRM-CM5	CLMcom-ETH-COSMO-crCLIM-v1-1	-2.2	-29.7	-0.1	-0.1	-1.1	0.3
CNRM-CERFACS-CNRM-CM5	CNRM-ALADIN63	0.5	1.2	-0.4	-0.2	-3.3	-0.2
CNRM-CERFACS-CNRM-CM5	DMI-HIRHAM5	-1.0	-20.2	-0.2	-0.1	-3.6	0.0
CNRM-CERFACS-CNRM-CM5	GERICS-REMO2015	-1.1	-27.8	-0.2	-0.3	-0.7	-0.1
CNRM-CERFACS-CNRM-CM5	ICTP-RegCM4-6	-0.3	-20.7	-0.3	-0.3	-3.4	0.2
CNRM-CERFACS-CNRM-CM5	KNMI-RACMO22E	-1.1	-17.4	0.2	0.0	1.4	-0.1
CNRM-CERFACS-CNRM-CM5	MOHC-HadREM3-GA7-05	1.5	4.1	-0.5	-0.7	-6.6	-1.4
CNRM-CERFACS-CNRM-CM5	RMIB-UGent-ALARO-0	-0.6		-0.3			
CNRM-CERFACS-CNRM-CM5	SMHI-RCA4	-1.0	-23.5	-0.1	-0.3	-1.0	0.0
ICHEC-EC-EARTH	CLMcom-CCLM4-8-17	-0.8	-16.3	-0.1	-0.2	0.7	0.0
ICHEC-EC-EARTH	CLMcom-ETH-COSMO-crCLIM-v1-1	-1.1	-13.7	0.0	-0.1	0.5	0.0
ICHEC-EC-EARTH	DMI-HIRHAM5	-0.8		0.1	0.2		0.0
ICHEC-EC-EARTH	ICTP-RegCM4-6	-0.6	-17.6	0.2	-0.1	2.1	-0.4
ICHEC-EC-EARTH	KNMI-RACMO22E	-0.8	-4.2	0.3	0.1	2.6	0.1
ICHEC-EC-EARTH	MOHC-HadREM3-GA7-05	1.6	15.9	-0.3	-0.6	-6.0	-1.5
ICHEC-EC-EARTH	SMHI-RCA4	0.0	-11.8	-0.1	-0.4	1.6	0.0
IPSL-IPSL-CM5A-MR	DMI-HIRHAM5	-2.5	-11.6	0.8	0.5	-0.6	0.2
IPSL-IPSL-CM5A-MR	GERICS-REMO2015	-2.2	-24.6	0.4	0.0	6.4	0.1
IPSL-IPSL-CM5A-MR	IPSL-INERIS-WRF331F	-3.3	-30.1	2.0	0.6		
IPSL-IPSL-CM5A-MR	KNMI-RACMO22E	-2.0	-1.5	0.9	0.4	2.6	0.3
IPSL-IPSL-CM5A-MR	SMHI-RCA4	-1.3	-10.8	0.5	0.0	2.5	0.5
MOHC-HadGEM2-ES	CLMcom-CCLM4-8-17	-2.4	-28.3	0.5	0.6	2.0	3.3
MOHC-HadGEM2-ES	CLMcom-ETH-COSMO-crCLIM-v1-1	-4.1	-35.2	0.8	0.9	6.2	1.5
MOHC-HadGEM2-ES	CNRM-ALADIN63	-2.5	-11.5	0.7	0.9	6.2	1.6
MOHC-HadGEM2-ES	DMI-HIRHAM5	-4.5	-30.0	0.9	1.1	5.3	1.9
MOHC-HadGEM2-ES	ICTP-RegCM4-6	-3.4	-35.0	0.8	1.0	5.8	1.8
MOHC-HadGEM2-ES	KNMI-RACMO22E	-3.5	-20.3	0.9	1.1	7.3	1.7
MOHC-HadGEM2-ES	MOHC-HadREM3-GA7-05	-1.3	-4.7	0.4	0.3	-0.3	0.2
MOHC-HadGEM2-ES	SMHI-RCA4	-3.2	-31.6	0.6	0.6	6.4	1.6
MPI-M-MPI-ESM-LR	CLMcom-CCLM4-8-17	-1.6	-20.2	0.1	0.2	6.4	1.1
MPI-M-MPI-ESM-LR	CLMcom-ETH-COSMO-crCLIM-v1-1	-2.0	-18.1	0.3	0.4	6.6	0.1
MPI-M-MPI-ESM-LR	CNRM-ALADIN63	0.0	-2.0	0.1	0.1	4.5	-0.5
MPI-M-MPI-ESM-LR	DMI-HIRHAM5	-1.6	-14.7	0.3	0.6	5.1	0.3
MPI-M-MPI-ESM-LR	GERICS-REMO2015	-2.0	-29.2	0.3	0.2	11.0	0.0
MPI-M-MPI-ESM-LR	ICTP-RegCM4-6	-1.5	-18.9	0.5	0.3	7.3	0.2
MPI-M-MPI-ESM-LR	KNMI-RACMO22E	-1.4	-6.0	0.5	0.5	8.4	0.2
MPI-M-MPI-ESM-LR	MOHC-HadREM3-GA7-05	0.3	7.7	0.2	0.0	1.6	-1.7
MPI-M-MPI-ESM-LR	MPI-CSC-REMO2009	-1.9	-26.3	0.4	0.3	10.7	-0.3
MPI-M-MPI-ESM-LR	SMHI-RCA4	-0.9	-17.6	0.2	0.0	8.5	0.0
NCC-NorESM1-M	CLMcom-ETH-COSMO-crCLIM-v1-1	-1.8	-22.0	-0.2	0.1	-0.4	1.3
NCC-NorESM1-M	CNRM-ALADIN63	0.4	0.0	-0.3	-0.1	-1.6	0.5
NCC-NorESM1-M	DMI-HIRHAM5	-2.1	-19.8	0.0	0.3	-0.9	1.9
NCC-NorESM1-M	GERICS-REMO2015	-1.5	-26.2	-0.1	0.0	1.2	0.9
NCC-NorESM1-M	ICTP-RegCM4-6	-0.9	-21.9	-0.1	0.0	-1.3	1.4
NCC-NorESM1-M	KNMI-RACMO22E	-1.3	-9.5	0.1	0.3	0.6	1.5
NCC-NorESM1-M	MOHC-HadREM3-GA7-05	0.9	5.6	-0.4	-0.3	-6.4	-0.2
NCC-NorESM1-M	SMHI-RCA4	-1.1	-21.1	-0.1	-0.1	-0.3	1.1

	TAS (K)	RSDS (W/m2)	PR (mm/day)	EVSPSBL (mm/day)	CLT (%)	PSL (hPa)
	ABS(0, 0.3)	ABS(0, 3)	ABS(0, 0.1)	ABS(0, 0.1)	ABS(0, 1)	ABS(0, 0.1)
	ABS(0.3, 0.6)	ABS(3, 6)	ABS(0.1, 0.3)	ABS(0.1, 0.3)	ABS(1, 3)	ABS(0.1, 0.3)
	ABS(0.6, 1.0)	ABS(6, 10)	ABS(0.3, 0.5)	ABS(0.3, 0.5)	ABS(3, 6)	ABS(0.3, 0.5)
	ABS(1.0, 1.5)	ABS(10, 15)	ABS(0.5, 0.8)	ABS(0.5, 0.8)	ABS(6, 9)	ABS(0.5, 0.8)
	ABS(>1.5)	ABS(>15)	ABS(>0.8)	ABS(>0.8)	ABS(>9)	ABS(>0.8)

Table. SM. 8: Mean summer discrepancies over continental Europe, computed over the red box (see Article, Fig. 2) for all EURO-CORDEX GCM/RCM pairs (detailed methods in Sect. 2.6).

GCM	RCM	TAS (K)	RSDS (W/m2)	PR (mm/day)	EVSPSBL (mm/day)	CLT (%)	PSL (hPa)
CNRM-CERFACS-CNRM-CM5	CLMcom-CCLM4-8-17	-0.8	-9.5	0.1	-0.1	0.6	-0.1
CNRM-CERFACS-CNRM-CM5	CLMcom-ETH-COSMO-crCLIM-v1-1	-0.6	-7.1	-0.1	-0.1	-1.4	-0.1
CNRM-CERFACS-CNRM-CM5	CNRM-ALADIN63	0.1	-2.7	0.2	0.1	-1.4	-0.3
CNRM-CERFACS-CNRM-CM5	DMI-HIRHAM5	-0.1	-4.9	0.0	-0.1	1.4	-0.1
CNRM-CERFACS-CNRM-CM5	GERICS-REMO2015	0.0	-4.6	0.0	-0.1	-0.7	-0.1
CNRM-CERFACS-CNRM-CM5	ICTP-RegCM4-6	-0.3	-4.4	0.0	-0.1	-0.2	0.6
CNRM-CERFACS-CNRM-CM5	KNMI-RACMO22E	0.3	-5.8	0.1	-0.1	-2.4	-0.2
CNRM-CERFACS-CNRM-CM5	MOHC-HadREM3-GA7-05	0.0	-1.3	0.0	0.0	-1.9	0.0
CNRM-CERFACS-CNRM-CM5	RMIB-UGent-ALARO-0	-1.1		0.0			
CNRM-CERFACS-CNRM-CM5	SMHI-RCA4	0.0	-8.5	0.2	0.0	0.6	0.3
ICHEC-EC-EARTH	CLMcom-CCLM4-8-17	-0.1	-3.5	0.0	-0.1	0.7	0.1
ICHEC-EC-EARTH	CLMcom-ETH-COSMO-crCLIM-v1-1	-0.3	-2.5	-0.1	-0.1	0.7	-0.4
ICHEC-EC-EARTH	DMI-HIRHAM5	0.2		-0.1	-0.1		-0.1
ICHEC-EC-EARTH	ICTP-RegCM4-6	0.2	0.2	0.0	-0.1	0.1	0.7
ICHEC-EC-EARTH	KNMI-RACMO22E	0.7	-0.9	0.0	-0.1	-1.0	-0.2
ICHEC-EC-EARTH	MOHC-HadREM3-GA7-05	0.6	2.9	0.0	0.0	-1.2	-0.3
ICHEC-EC-EARTH	SMHI-RCA4	0.7	-3.2	0.0	0.0	1.0	0.3
IPSL-IPSL-CM5A-MR	DMI-HIRHAM5	0.2	-2.5	0.3	0.1	4.6	-0.1
IPSL-IPSL-CM5A-MR	GERICS-REMO2015	0.1	-3.3	0.2	0.1	4.2	-0.4
IPSL-IPSL-CM5A-MR	IPSL-INERIS-WRF331F	0.7	-7.9	0.3	0.1		
IPSL-IPSL-CM5A-MR	KNMI-RACMO22E	0.8	-4.0	0.3	0.1	1.3	-0.2
IPSL-IPSL-CM5A-MR	SMHI-RCA4	0.2	-5.2	0.4	0.3	3.7	0.5
MOHC-HadGEM2-ES	CLMcom-CCLM4-8-17	-0.9	-7.3	0.0	-0.1	2.2	1.1
MOHC-HadGEM2-ES	CLMcom-ETH-COSMO-crCLIM-v1-1	-1.2	-5.4	-0.1	-0.2	1.5	0.7
MOHC-HadGEM2-ES	CNRM-ALADIN63	-0.6	-2.5	0.2	0.0	2.0	-0.2
MOHC-HadGEM2-ES	DMI-HIRHAM5	-1.1	-2.7	0.0	-0.2	4.9	0.4
MOHC-HadGEM2-ES	ICTP-RegCM4-6	-1.2	-1.8	-0.1	-0.1	0.9	0.7
MOHC-HadGEM2-ES	KNMI-RACMO22E	0.1	-3.8	0.1	-0.1	0.3	0.4
MOHC-HadGEM2-ES	MOHC-HadREM3-GA7-05	-0.4	0.7	0.1	0.0	-0.3	0.3
MOHC-HadGEM2-ES	SMHI-RCA4	-0.7	-5.5	0.0	0.0	3.3	0.8
MPI-M-MPI-ESM-LR	CLMcom-CCLM4-8-17	-0.2	-5.3	0.2	0.1	1.7	0.2
MPI-M-MPI-ESM-LR	CLMcom-ETH-COSMO-crCLIM-v1-1	-0.3	-4.4	0.2	0.0	2.8	-1.1
MPI-M-MPI-ESM-LR	CNRM-ALADIN63	0.2	-0.6	0.4	0.2	2.4	-1.2
MPI-M-MPI-ESM-LR	DMI-HIRHAM5	0.2	-1.9	0.2	0.1	3.4	-0.5
MPI-M-MPI-ESM-LR	GERICS-REMO2015	0.3	-1.1	0.1	0.0	1.0	-0.1
MPI-M-MPI-ESM-LR	ICTP-RegCM4-6	0.5	-2.0	0.3	0.1	1.5	-0.4
MPI-M-MPI-ESM-LR	KNMI-RACMO22E	0.8	-3.3	0.2	0.1	1.5	-0.9
MPI-M-MPI-ESM-LR	MOHC-HadREM3-GA7-05	0.3	0.8	0.3	0.2	0.9	-1.3
MPI-M-MPI-ESM-LR	MPI-CSC-REMO2009	0.3	-2.7	0.2	0.1	2.8	-1.1
MPI-M-MPI-ESM-LR	SMHI-RCA4	0.6	-4.7	0.4	0.2	4.2	-0.5
NCC-NorESM1-M	CLMcom-ETH-COSMO-crCLIM-v1-1	-0.8	-9.8	0.0	-0.1	2.5	-0.8
NCC-NorESM1-M	CNRM-ALADIN63	-0.3	-3.4	0.3	0.1	1.2	-1.0
NCC-NorESM1-M	DMI-HIRHAM5	-0.7	-7.3	0.1	-0.1	3.4	-0.2
NCC-NorESM1-M	GERICS-REMO2015	-0.1	-7.2	0.1	-0.1	2.0	-0.6
NCC-NorESM1-M	ICTP-RegCM4-6	-0.2	-6.9	0.1	0.0	1.3	0.2
NCC-NorESM1-M	KNMI-RACMO22E	0.0	-7.3	0.1	0.0	0.8	-0.6
NCC-NorESM1-M	MOHC-HadREM3-GA7-05	-0.1	-3.3	0.1	0.0	-0.3	-0.5
NCC-NorESM1-M	SMHI-RCA4	0.0	-9.5	0.1	0.1	2.3	-0.3

	TAS (K)	RSDS (W/m2)	PR (mm/day)	EVSPSBL (mm/day)	CLT (%)	PSL (hPa)
	ABS(0, 0.3)	ABS(0, 3)	ABS(0, 0.1)	ABS(0, 0.1)	ABS(0, 1)	ABS(0, 0.1)
	ABS(0.3, 0.6)	ABS(3, 6)	ABS(0.1, 0.3)	ABS(0.1, 0.3)	ABS(1, 3)	ABS(0.1, 0.3)
	ABS(0.6, 1.0)	ABS(6, 10)	ABS(0.3, 0.5)	ABS(0.3, 0.5)	ABS(3, 6)	ABS(0.3, 0.5)
	ABS(1.0, 1.5)	ABS(10, 15)	ABS(0.5, 0.8)	ABS(0.5, 0.8)	ABS(6, 9)	ABS(0.5, 0.8)
	ABS(>1.5)	ABS(>15)	ABS(>0.8)	ABS(>0.8)	ABS(>9)	ABS(>0.8)

Table. SM. 9: Same as Table SM. 8, but for winter.

REF/ALD inconsistencies for each experiment

PAIR	TAS (JJA)	RSDS (JJA)	PR (JJA)	EVSPSBL (JJA)	CLT (JJA)	PSL (JJA)	MEAN (JJA)
REF/ALD150	50.6	65.3	35.4	50.9	45.2	25.5	45.5

GCM/RCM	TAS (DJF)	RSDS (DJF)	PR (DJF)	EVSPSBL (DJF)	CLT (DJF)	PSL (DJF)	MEAN (DJF)
REF/ALD150	43.7	57.7	32.8	46.3	40.6	23.6	40.8

Value
[0,20)
[20,40)
[40,60)
[60,80)
[80,100]

Table. SM. 10: Fraction of incoherent grid points computed over land part of the standard EUR-11 domain (black box in Article, Fig. 2) for the benchmark experiment REF/ALD150 (see methods 2.7) using historical seasonal discrepancies (1970-1999).

Model	TAS (JJA)	RSDS (JJA)	PR (JJA)	EVSPSBL (JJA)	CLT (JJA)	PSL (JJA)	Mean (JJA)
ALD150	10.5	12.0	26.6	35.1	7.4	6.8	16.4
ALD150CO2	19.0	7.2	17.0	39.5	5.2	21.0	18.1
ALD12	22.3	29.3	34.3	44.1	24.2	10.3	27.4
ALD150AERO	39.5	96.9	23.2	35.2	10.1	41.0	41.0
ALD150PHYS	20.8	58.3	41.5	57.7	46.7	41.0	44.3
ALD12MAX	76.2	100.0	52.2	60.5	45.2	61.6	66.0

Model	TAS (DJF)	RSDS (DJF)	PR (PR)	EVSPSBL (DJF)	CLT (DJF)	PSL (DJF)	Mean (DJF)
ALD150CO2	20.7	9.8	12.7	35.4	14.9	0.4	15.7
ALD150	20.7	14.2	15.7	28.6	17.3	22.7	19.9
ALD12	53.1	22.0	14.2	53.1	14.9	10.3	28.0
ALD150PHYS	41.3	72.7	18.6	44.3	42.8	14.9	39.1
ALD150AERO	44.6	84.5	14.6	56.6	17.5	33.2	41.9
ALD12MAX	51.7	71.0	14.9	69.0	46.3	11.8	44.1

Value
[0,20)
[20,40)
[40,60)
[60,80)
[80,100]

Table. SM. 11: Fraction of incoherent grid points computed over land part of the standard EUR-11 domain (black box in Article, Fig. 2) for each experiment (see methods 2.7) using the seasonal discrepancies in projected climate change between RCP8.5 (2070-2099) and historical (1970-1999) periods. The fraction is calculated as the ratio between the number of land points where the differences in projected change for a given variable between REF and ALD are statistically significant and the total number of land points in the EUR-11 domain.

	ALD150 (JJA)	ALD12 (JJA)	ALD150AERO (JJA)	ALD150CO2 (JJA)	ALD150PHYS (JJA)	ALD12MAX (JJA)
TAS (K)	0.1	-0.1	-0.8	0.0	0.0	-1.3
RSDS (W/m2)	1.5	-1.5	-19.3	0.7	-2.4	-31.1
PR (mm/day)	-0.2	0.3	-0.1	0.0	0.6	0.5
EVSPSBL (mm/day)	-0.1	0.2	-0.1	0.1	0.4	0.2
CLT (%)	-0.9	0.0	-1.7	-0.8	-0.5	1.1
PSL (hPa)	0.3	0.0	0.5	0.3	0.0	0.3

	ALD150 (DJF)	ALD12 (DJF)	ALD150AERO (DJF)	ALD150CO2 (DJF)	ALD150PHYS (DJF)	ALD12MAX (DJF)
TAS (K)	-0.1	0.5	-0.4	-0.2	0.3	0.3
RSDS (W/m2)	0.2	-0.2	-4.3	-0.1	1.3	-3.9
PR (mm/day)	0.0	0.1	0.0	0.0	-0.1	0.0
EVSPSBL (mm/day)	0.0	0.0	-0.1	0.0	0.0	0.0
CLT (%)	-0.6	-0.5	0.2	-0.5	0.6	1.4
PSL (hPa)	0.8	0.4	0.8	0.5	0.6	0.6

	TAS (K)	RSDS (W/m2)	PR (mm/day)	EVSPSBL (mm/day)	CLT (%)	PSL (hPa)
	ABS(0, 0.3)	ABS(0, 3)	ABS(0, 0.1)	ABS(0, 0.1)	ABS(0, 1)	ABS(0, 0.1)
	ABS(0.3, 0.6)	ABS(3, 6)	ABS(0.1, 0.3)	ABS(0.1, 0.3)	ABS(1, 3)	ABS(0.1, 0.3)
	ABS(0.6, 1.0)	ABS(6, 10)	ABS(0.3, 0.5)	ABS(0.3, 0.5)	ABS(3, 6)	ABS(0.3, 0.5)
	ABS(1.0, 1.5)	ABS(10, 15)	ABS(0.5, 0.8)	ABS(0.5, 0.8)	ABS(6, 9)	ABS(0.5, 0.8)
	ABS(>1.5)	ABS(>15)	ABS(>0.8)	ABS(>0.8)	ABS(>9)	ABS(>0.8)

Table. SM. 12: Mean seasonal discrepancies over continental Europe, computed over the red box (see Article, Fig. 2) for all experiments (detailed methods in Sect. 2.6).